Distribution of Segment Lengths in Genome Rearrangements

Glenn Tesler*

Department of Mathematics
University of California, San Diego, USA
gptesler@math.ucsd.edu

Submitted: Nov 13, 2007; Accepted: Aug 3, 2008; Published: Aug 11, 2008
Mathematics Subject Classifications: 05A15, 92D15, 92D20

Abstract

The study of gene orders for constructing phylogenetic trees was introduced by Dobzhansky and Sturtevant in 1938. Different genomes may have homologous genes arranged in different orders. In the early 1990s, Sankoff and colleagues modelled this as ordinary (unsigned) permutations on a set of numbered genes $1, 2, \ldots, n$, with biological events such as inversions modelled as operations on the permutations. Signed permutations may be used when the relative strands of the genes are known, and "circular permutations" may be used for circular genomes. We use combinatorial methods (generating functions, commutative and noncommutative formal power series, asymptotics, recursions, and enumeration formulas) to study the distributions of the number and lengths of conserved segments of genes between two or more unichromosomal genomes, including signed and unsigned genomes, and linear and circular genomes. This generalizes classical work on permutations from the 1940s–60s by Wolfowitz, Kaplansky, Riordan, Abramson, and Moser, who studied decompositions of permutations into strips of ascending or descending consecutive numbers. In our setting, their work corresponds to comparison of two unsigned genomes (known gene orders, unknown gene orientations). Maple software implementing our formulas is available at `http://www.math.ucsd.edu/~gptesler/strips` .

1 Introduction

The study of gene orders in phylogenetics was introduced by Dobzhansky and Sturtevant, 1938 [11], in a study of inversions in *Drosophila pseudoobscura*. More recently, in the late 1980s, Jeffrey Palmer and colleagues [21, 22] compared the mitochondiral genomes of

*Funded by a Sloan Research Fellowship in Molecular Biology and NSF Grant DMS-0718810. The author also thanks the anonymous referee for helpful suggestions on presentation.

cabbage and turnip, and found that the DNA sequences of many genes are more than 99% identical. However, the order of the genes was quite different. These and similar studies have shown that genome rearrangements are an important form of molecular evolution.

To study genome rearrangements, conserved segments between two genomes must be identified. Traditionally, this has been done by identifying homologous genes between the genomes, and determining runs of genes that are consecutive in both genomes. The pre-sequencing era methods for identifying the locations (and hence order) of the genes include inference from linkage maps and recombination rates [20] and radiation hybrid maps [9, 19]. These methods do not identify on which of the two strands a gene is located. Thus, these methods give the gene order in one genome as an unsigned permutation of the gene order in the other genome (when both have one chromosome; the multichromosomal situation is similar but involves partitioning the permutation). The relative orientation of a *singleton* segment (a conserved segment containing one gene) cannot be determined. When a segment with 2 or more genes has the same genes in the same order in both genomes, it is inferred that the corresponding genes have the same orientations in both genomes, while if they run in the exact opposite order, it is inferred that they have opposite orientations. It is possible that individual genes have been flipped, but this cannot be detected. Sampling the genes with the same methodology at a higher resolution might resolve this partially but will ultimately just push the problem of misclassified orientations to a finer level of resolution rather than solve it.

More recently, as the DNA sequences of various genomes have become available, determination of homologous genes and of conserved segments has been done by comparison of the DNA sequences. This allows a more precise determination of the coordinates of each common feature, as well as its orientation (one of two strands). Thus, sequence comparison gives the gene or segment order in one genome as a signed permutation of the order in the other genome, when both have one chromosome (again, this can be extended to multiple chromosomes). It is convenient to consecutively label the elements of the "reference" genome $1, \ldots, n$ in the linear order in which they appear, and to describe the second genome as a permutation of those labels.

The numbers $1, \ldots, n$ represent homologous markers, whether based on genes or aligned sequences. If signed permutations are used, the signs represent their strand.

The simplest type of genome rearrangement, known as an *inversion* or *reversal*, takes a segment of consecutive genes and reverses their order, and in the signed case, additionally inverts their signs. See Figure 1. Reversals (and other genome rearrangements) disrupt runs of consecutive elements, breaking them into multiple runs, which we call *strips*.

In this paper, we will consider the problem of decomposing unsigned permutations of $1, \ldots, n$ into ascending strips $i, i+1, \ldots, j$ or descending strips $j, j-1, \ldots, i$, and decomposing signed permutations of $1, \ldots, n$ into ascending strips $i, i+1, \ldots, j$ or descending strips $-j, -(j-1), \ldots, -i$; for descending unsigned strips, $0 < i < j < n$, and for the others, $0 < i \leq j < n$. The strips represent conserved segments. We will count the number of signed or unsigned permutations of $1, \ldots, n$ that decompose into k strips. More generally, we will handle multiple genomes, circular genomes, and the lengths of the strips.

Further extensions of this, which we do not treat in this paper, could be to genomes

(a) Unsigned rearrangements

$$\begin{array}{l}
1\ \underline{2\ 3\ 4\ 5\ 6\ 7}\ 8\ 9\\
1\ 7\ 6\ \underline{5\ 4\ 3\ 2\ 8}\ 9\\
1\ 7\ 6\ 8\ 2\ 3\ \underline{4}\ 5\ 9\\
1\ 7\ 6\ 8\ 2\ 3\ 4\ 5\ 9
\end{array}$$

(b) Signed rearrangements

$$\begin{array}{l}
1\ \underline{2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7}\ 8\ 9\\
1\ -7\ -6\ \underline{-5\ -4\ -3\ -2\ 8}\ 9\\
1\ -7\ -6\ -8\ 2\ 3\ \underline{4}\ 5\ 9\\
1\ -7\ -6\ -8\ 2\ 3\ -4\ 5\ 9
\end{array}$$

(c) Unsigned arrangement

$\sigma^{(1)}:\ \langle 1,2,3,4,5,6,7,8,9\rangle$
$\sigma^{(2)}:\ \langle 1,7,6,8,2,3,4,5,9\rangle$

(d) Signed arrangement

$\sigma^{(1)}:\ \langle 1,2,3,4,5,6,7,8,9\rangle$
$\sigma^{(2)}:\ \langle 1,-7,-6,-8,2,3,-4,5,9\rangle$

(e) Unsigned strips

$\sigma^{(1)}:\ \boxed{1},\boxed{2,3,4,5},\boxed{6,7},\boxed{8},\boxed{9}$
$\sigma^{(2)}:\ \boxed{1},\boxed{7,6},\boxed{8},\boxed{2,3,4,5},\boxed{9}$

(f) Signed strips

$\sigma^{(1)}:\ \boxed{1},\boxed{2,3},\boxed{4},\boxed{5},\boxed{6,7},\boxed{8},\boxed{9}$
$\sigma^{(2)}:\ \boxed{1},\boxed{-7,-6},\boxed{-8},\boxed{2,3},\boxed{-4},\boxed{5},\boxed{9}$

Figure 1: (a,b) A sequence of 3 reversals applied to the identity permutation. In the unsigned case, the order of elements in the underlined segment is reversed. In the signed case, the order is reversed and the signs are inverted. (c,d) Comparing just the first and last permutation in each scenario gives (un)signed $(9,2)$-arrangements (9 genes, 2 genomes). (e,f) Strips (preserved intervals) in these arrangements have ordered types $(1,4,2,1,1)$ (unsigned) and $(1,2,1,1,2,1,1)$ (signed), by listing the lengths of consecutive strips in $\sigma^{(1)}$. The unordered types are $(4,2,1,1,1)$ (unsigned) and $(2,2,1,1,1,1,1)$ (signed).

with multiple chromosomes; genomes with *equal content* repeats (each value $i=1,\ldots,n$ appears the same number of times in all genomes, counting both $\pm i$ equivalently); and genomes with *unequal content* (the multiplicity of a gene varies from genome to genome).

We have written Maple software that implements our formulas. In addition, for small numbers of genes and genomes, we include a program to list all unsigned arrangements and analyze the strip lengths, to compare with the counts and generating functions given by the formulas. The software is available at `http://www.math.ucsd.edu/~gptesler/strips`.

Counting strips in two unsigned permutations is equivalent to a problem treated in a series of papers from the 1940s–60s, that consider the number of unsigned permutations on $1,\ldots,n$ with exactly t pairs of adjacent positions of the form $i,i{+}1$ or $i{+}1,i$. In our setting, this is the same as having exactly $k=n-t$ unsigned strips. Wolfowitz, 1942 [33, Sections 6–7] initiated these studies. Wolfowitz, 1944 [34] gave an asymptotic formula; Kaplansky, 1945 [15] gave two additional subdominant terms of the asymptotic formula; Riordan, 1965 [28] gave a generating function and a recurrence equation. Abramson and Moser, 1967 [1] gave an explicit multiple summation formula for the number of permutations of $1,\ldots,n$ with exactly k strips and various conditions on the lengths of the strips. This paper generalizes all of these to signed permutations and to multiple genomes.

The model of conserved segments as strips is idealized. Recent papers that treat higher resolution data use *syntenic blocks* in place of conserved segments. These blocks ignore minor perturbations in gene order that occur below a specified resolution; this effectively merges several strips into one block. Pevzner and Tesler, 2003 [25] introduced the first algorithm to construct syntenic blocks that explicitly took such small scale rearrangements into account. This was for high resolution data from genome alignments, which may be regarded as signed permutations. Murphy et al., 2005 [19] used a different algorithm adapted to radiation hybrid maps, which may be regarded as unsigned permutations.

In Section 2, we introduce notation for multiple genome arrangements and give examples of breaking a three genome arrangement into strips, in several variations (signed or unsigned genomes; ordered or unordered types and weights). We also give basic results on compressing an arrangement by collapsing each strip into a single number.

In Section 3, we develop formulas to enumerate signed arrangements by ordered and unordered types, and in Section 4, we develop generating functions for ordered types. We also count arrangements by number of strips, count incompressible arrangements (all strip lengths equal 1), and give asymptotic formulas. Then in Section 5, we use formal power series to establish a relationship between the unsigned and signed cases, and use that relationship to develop formulas for enumeration of unsigned arrangements by ordered types. Section 6 gives generating functions (signed and unsigned cases) for unordered types. Section 7 gives a worked out example of these computations. Section 9 extends all this to circular genomes.

In Section 8, we also consider ramifications in genome studies: issues in signed vs. unsigned data; quantifying an error in Sankoff and Trinh [29, 30]; imposing a minimum or maximum length on strips; and issues in incompressible permutations;

In Section 10, we compute the mean and variance of the number of strips over all arrangements. In Section 11, we develop recursions and mixed recursions / differential equations that provide an alternate means to compute generating functions and counts.

Some proofs are delayed to Appendix A.

2 Introductory example and notation

Let $\mathcal{S}_n$ denote the set of permutations on $1,\ldots,n$ and $\mathcal{B}_n$ denote the set of signed permutations on $1,\ldots,n$. We use *one-line form,* e.g., $\langle 1,3,4,2\rangle \in \mathcal{S}_4$ and $\langle 1,-3,4,-2\rangle \in \mathcal{B}_4$. In this notation, the *identity permutation of length* n is $\mathrm{id}_n = \langle 1,\ldots,n\rangle$.

We consider $g \geq 2$ genomes at a time. An *unsigned* (n,g)*-arrangement* is a g-tuple $\vec{\sigma} = (\sigma^{(1)},\ldots,\sigma^{(g)})$ of permutations in $\mathcal{S}_n$ where $\sigma^{(1)} = \mathrm{id}_n$. (We consecutively label the elements of the first genome $1,\ldots,n$, and represent the other genomes as permutations of that.) $\mathcal{A}_n^{(g)}$ is the set of all unsigned (n,g)-arrangements and $\mathcal{A}^{(g)} = \cup_{n=0}^{\infty}\mathcal{A}_n^{(g)}$ is unsigned arrangements of all sizes on g genomes.

A *signed* (n,g)*-arrangement* is a g-tuple $\vec{\sigma} = (\sigma^{(1)},\ldots,\sigma^{(g)})$ of permutations in $\mathcal{B}_n$ where $\sigma^{(1)} = \mathrm{id}_n$. $\mathcal{B}_n^{(g)}$ is the set of all signed (n,g)-arrangements, and $\mathcal{B}^{(g)} = \cup_{n=0}^{\infty}\mathcal{B}_n^{(g)}$ is signed arrangements of all sizes on g genomes. See Table 1 for a summary of notation.

In an unsigned (n,g)-arrangement, consecutive entries (i,j) of $\sigma^{(1)}$ form an *adjacency* if i,j or j,i are consecutive in each of $\sigma^{(2)},\sigma^{(3)},\ldots$; otherwise, (i,j) (and (j,i)) is a *breakpoint* of $\sigma^{(1)}$. In a signed (n,g)-arrangement, consecutive entries (i,j) of $\sigma^{(1)}$ form an *adjacency* if i,j or $-j,-i$ are consecutive in each of $\sigma^{(2)},\sigma^{(3)},\ldots$; otherwise, (i,j)

Description	Symbol
Identity permutation of size n	$\mathrm{id}_n = \langle 1, \ldots, n \rangle$
Arrangement with g genomes	$\vec{\sigma} = (\sigma^{(1)}, \ldots, \sigma^{(g)})$, with $\sigma^{(1)} = \mathrm{id}_n$
	Also $\vec{\pi}$ (unsigned), $\vec{\tau}$ (compressed)
Vector of g positive signs	$\vec{\epsilon}_+ = (+1, \ldots, +1)$ (len. g)
# sign vectors $\neq \vec{\epsilon}_+$	$G = 2^{g-1} - 1$. Also define $\widetilde{G} = 2^{1-g} - 1$.
Length of permutation/composition	$\ell(\mu)$
# parts equal i	$m_i(\mu)$
# permutations of partition μ	$M(\mu) = \ell(\mu)!/(m_1(\mu)!\, m_2(\mu)! \ldots)$
Map from signed to unsigned weights	$\phi(f)$, has inverse ϕ^{-1}

Description	Ordered types	Unordered types
Set of types for size n	Compositions: $\mathcal{C}_n$	Partitions: $\mathcal{P}_n$
with k nonzero parts	$\mathcal{C}_{n,k}$	$\mathcal{P}_{n,k}$

Description	Unsigned arrangements	Signed arrangements
Set of permutations of size n	$\mathcal{S}_n$	$\mathcal{B}_n$
Set of arr. on g genomes	$\mathcal{A}^{(g)}$	$\mathcal{B}^{(g)}$
with n elements	$\mathcal{A}^{(g)}_n$	$\mathcal{B}^{(g)}_n$
and k strips	$\mathcal{A}^{(g)}_{n,k}$	$\mathcal{B}^{(g)}_{n,k}$
# (n,g)-arrs. with k strips	$a^{(g)}_{n,k} = \lvert\mathcal{A}^{(g)}_{n,k}\rvert$	$b^{(g)}_{n,k} = \lvert\mathcal{B}^{(g)}_{n,k}\rvert$
ogf for fixed n, varying k	$a^{(g)}_n(z) = \sum_{k=0}^{n} a^{(g)}_{n,k} z^n$	$b^{(g)}_n(z) = \sum_{k=0}^{n} b^{(g)}_{n,k} z^n$
ogf for varying n, k	$a^{(g)}(t,z) = \sum_{n=0}^{\infty} t^n\, a^{(g)}_n(z)$	$b^{(g)}(t,z) = \sum_{n=0}^{\infty} t^n\, b^{(g)}_n(z)$

Description	Unsigned ordered	Unsigned unordered	Signed ordered	Signed unordered
Type of an arr.	$\alpha \in \mathcal{C}_n$	$\lambda \in \mathcal{P}_n$	$\beta \in \mathcal{C}_n$	$\mu \in \mathcal{P}_n$
# arrs. by type	$A^{(g)}_\alpha$	$a^{(g)}_\lambda$	$B^{(g)}_\beta$	$b^{(g)}_\mu$
Wt. of length n strip	U_n	u_n	V_n	v_n
ogf	$U(t) = \sum_{n=1}^{\infty} t^n U_n$	$u(t)$	$V(t)$	$v(t)$
vector	$\vec{U} = (U_1, U_2, \ldots)$	$\vec{u}$	$\vec{V}$	$\vec{v}$
Wt. of one arr.	$U_\alpha = U_{\alpha_1} U_{\alpha_2} \cdots$	u_λ	V_β	v_μ
Wt. of set S of arrs.	$\omega_A(S)$	$\omega_a(S)$	$\omega_B(S)$	$\omega_b(S)$
Wt. of all (n,g)-arrs.	$A^{(g)}_n(\vec{U}) = \sum_\alpha A^{(g)}_\alpha U_\alpha$	$a^{(g)}_n(\vec{u})$	$B^{(g)}_n(\vec{V})$	$b^{(g)}_n(\vec{v})$
Wt. over all n	$A^{(g)}(\vec{U};t) = \sum_{n=0}^{\infty} t^n A^{(g)}_n(\vec{U})$	$a^{(g)}(\vec{u};t)$	$B^{(g)}(\vec{V};t)$	$b^{(g)}(\vec{v};t)$

Table 1: Summary of notation for linear arrangements. When a formula is given in only one column, use a similar formula in the other columns, substituting the corresponding notation for each column. Abbreviations: "arr(s)." is arrangement(s); "wt." is weight; "ogf" is ordinary generating function.

(and $(-j,-i)$) is a *breakpoint* of $\sigma^{(1)}$. Since we always set $\sigma^{(1)} = \langle 1,\ldots,n\rangle$ in this paper, consecutive entries in $\sigma^{(1)}$ have the form $(j-1,j)$ in both the unsigned and signed cases.

Watterson et al., 1982 [32] used breakpoints for two unsigned unichromosomal circular genomes, using a symbolic representation of gene orders. Formal definitions for unsigned permutations were given by Kececioglu and Sankoff, 1993 [16, 18] and Bafna and Pevzner, 1993 [5, 6], and for signed permutations by [5, 6] and Kececioglu and Sankoff, 1994 [17]. Hannenhalli and Pevzner, 1995 [12] generalized it to two genomes with multiple chromosomes, and Tesler and Pevzner, 2003 [26] made further definitions about the chromosome ends. Our notion of breakpoints corresponds to *internal breakpoints* in [26]; we do not count *external breakpoints* at the ends of the chromosomes (when the first entries are not all the same, or the last entries are not all the same).

A *strip* is a sequence of consecutive entries of $\sigma^{(1)}$ terminated on both sides either by the start/end of the permutation, or a breakpoint. For $n \geq 1$, the number of strips is one more than the number of breakpoints. For $n = 0$, there is a unique arrangement (the *null arrangement*) and it has 0 strips. A *singleton* is a strip of length 1.

Let $a^{(g)}_{n,k}$ be the number of unsigned (n,g)-arrangements that break into k strips, and $b^{(g)}_{n,k}$ be the number of signed (n,g)-arrangements that break into k strips.

Example 2.1. Consider these signed permutations (in *one-line notation*):

$$\begin{array}{ll}\sigma^{(1)}: & \langle 1,2,3,4,5,6,7,8,9,10,11,12,13\rangle\\ \sigma^{(2)}: & \langle -9,8,-7,-6,-5,10,11,12,1,2,3,4,-13\rangle\\ \sigma^{(3)}: & \langle -4,-3,-2,-1,5,6,7,8,9,10,11,12,13\rangle\end{array}$$

There are $g = 3$ signed permutations, each on $n = 13$ elements, and $\vec{\sigma} = (\sigma^{(1)},\sigma^{(2)},\sigma^{(3)})$ is a *signed* $(13,3)$*-arrangement.*

The are 5 breakpoints in $\sigma^{(1)}$: $(4,5),(7,8),(8,9),(9,10),(12,13)$. This breaks this arrangement into $k = 5+1 = 6$ strips:

$$\begin{array}{ll}\sigma^{(1)}: & \boxed{1,2,3,4}, \boxed{5,6,7}, \boxed{8}, \boxed{9}, \boxed{10,11,12}, \boxed{13}\\ \sigma^{(2)}: & \boxed{-9}, \boxed{8}, \boxed{-7,-6,-5}, \boxed{10,11,12}, \boxed{1,2,3,4}, \boxed{-13}\\ \sigma^{(3)}: & \boxed{-4,-3,-2,-1}, \boxed{5,6,7}, \boxed{8}, \boxed{9}, \boxed{10,11,12}, \boxed{13}\end{array}$$

.

The *ordered type* of this arrangement is the lengths of the consecutive strips in $\sigma^{(1)}$: $\beta = (4,3,1,1,3,1)$. It is a *composition* of n: $13 = 4+3+1+1+3+1$ is expressed as a sum of positive integers. Let $\mathcal{C}_n$ denote the set of all compositions of n and $\mathcal{C}_{n,k}$ denote the set of all compositions of n into exactly k nonzero parts. For $n > 0$, $|\mathcal{C}_n| = 2^{n-1}$ and for $n \geq k > 0$, $|\mathcal{C}_{n,k}| = \binom{n-1}{k-1}$ while $|\mathcal{C}_{n,0}| = 0$. For $n = 0$, there is a null composition, so $|\mathcal{C}_0| = |\mathcal{C}_{0,0}| = 1$ while $|\mathcal{C}_{0,k}| = 0$ for $k > 0$.

We may also consider the *unordered type* of this arrangement, which is the lengths of the strips listed in decreasing order $\mu = (4,3,3,1,1,1)$. This is a *partition* of n: $13 = 4+3+3+1+1+1$ is expressed as a sum of weakly decreasing positive integers. Let $\mathcal{P}_n$ denote the set of all partitions of n and $\mathcal{P}_{n,k}$ denote the set of all partitions of n into exactly k nonzero parts. The cardinalities of these sets, $p(n) = |\mathcal{P}_n|$ and $p(n,k) = |\mathcal{P}_{n,k}|$, have been studied extensively for centuries; for surveys, see Dickson, 1920 [10, Ch. 3], Andrews, 1976 [2], and Andrews and Eriksson, 2004 [3].

The *ordered weight* of this arrangement is $V_4V_3V_1V_1V_3V_1$, where the V_i's are noncommuting variables. The *unordered weight* is $v_4{v_3}^2{v_1}^3$, where the v_i's are commuting variables. The (un)ordered weight of a set of arrangements is the sum of the weights of the arrangements in the set. We will compute generating functions for the weights of all arrangements, subclassified in various ways.

Note that if the second or third genome were used as the reference instead of the first, the ordered type and weight would change (since the strips would be in a different left-to-right order) but the unordered type and weight would not change.

For a partition or composition μ, let $\ell(\mu)$ be the number of nonzero parts and $m_i(\mu)$ be the number of parts equal to i (for $i > 0$). When we use unordered types (partitions), many different ordered types (compositions) are combined; specifically, for a partition μ, the number of distinct compositions obtained by permuting its nonzero parts is

$$M(\mu) = \binom{\ell(\mu)}{m_1(\mu), m_2(\mu), \ldots} = \frac{\ell(\mu)!}{m_1(\mu)!\, m_2(\mu)! \cdots}\,.$$

The *strips* in this arrangement are $J_1 = \langle 1,2,3,4\rangle$, $J_2 = \langle 5,6,7\rangle$, $J_3 = \langle 8\rangle$, $J_4 = \langle 9\rangle$, $J_5 = \langle 10,11,12\rangle$, $J_6 = \langle 13\rangle$. The *negative* of strip $J = \langle j_1, j_2, \ldots, j_m\rangle$ is $-J = \langle -j_m, \ldots, -j_2, -j_1\rangle$, while its *reverse* is $J^r = \langle j_m, \ldots, j_2, j_1\rangle$.

The representation of $\vec{\sigma}$ in terms of concatenations of these strips is

$$\begin{aligned}
\sigma^{(1)} &: \ J_1, J_2, J_3, J_4, J_5, J_6 \\
\sigma^{(2)} &: \ -J_4, J_3, -J_2, J_5, J_1, -J_6 \\
\sigma^{(3)} &: \ -J_1, J_2, J_3, J_4, J_5, J_6
\end{aligned}$$

The *(signed) compression* of $\vec{\sigma}$ is obtained by replacing $\pm J_i$ with $\pm i$:

$$\begin{aligned}
\tau^{(1)} &: \ \langle 1,2,3,4,5,6\rangle \\
\tau^{(2)} &: \ \langle -4,3,-2,5,1,-6\rangle \\
\tau^{(3)} &: \ \langle -1,2,3,4,5,6\rangle
\end{aligned}$$

A signed (n,g)-arrangement is *incompressible* if it equals its compression. This is equivalent to any of these conditions: it has no adjacencies; all its strips are singletons; its type has form (1^n). Note that the compression of a signed (n,g)-arrangement is incompressible.

Let $\mathcal{B}^{(g)}_{n,k}$ be the subset of $\mathcal{B}^{(g)}_n$ consisting of signed (n,g)-arrangements that break into k strips, and $b^{(g)}_{n,k} = |\mathcal{B}^{(g)}_{n,k}|$ be the number of such arrangements. Note that $\mathcal{B}^{(g)}_{n,n}$ is the set of incompressible signed (n,g)-arrangements. With this notation, the example above illustrates the following:

Theorem 2.2. *The procedure illustrated above gives a bijection*

$$\Psi_b \ : \ \mathcal{B}^{(g)}_{n,k} \to \mathcal{B}^{(g)}_{k,k} \times \mathcal{C}_{n,k}$$

between signed (n,g)-arrangements with k strips and ordered pairs $(\vec{\tau}, \beta)$ where *(i)* $\vec{\tau} = (\tau^{(1)}, \ldots, \tau^{(g)}) \in \mathcal{B}_k^{(g)}$ *is incompressible;*

(ii) $\beta \in \mathcal{C}_{n,k}$ *is the ordered type of the arrangement.*

Example 2.3. Here is a similar example with unsigned permutations, obtained by dropping the signs in Example 2.1. Let $\vec{\pi} = |\vec{\sigma}|$ where $\vec{\sigma}$ is given in Example 2.1 and $|\vec{\sigma}|$ denotes taking the absolute value of all elements in each of $\sigma^{(1)}, \ldots, \sigma^{(g)}$:

$$\begin{array}{ll} \pi^{(1)}: & \langle 1,2,3,4,5,6,7,8,9,10,11,12,13\rangle \\ \pi^{(2)}: & \langle 9,8,7,6,5,10,11,12,1,2,3,4,13\rangle \\ \pi^{(3)}: & \langle 4,3,2,1,5,6,7,8,9,10,11,12,13\rangle \end{array}$$

This breaks into $k = 4$ unsigned strips:

$$\begin{array}{llll} \pi^{(1)}: & \boxed{1,2,3,4}, \boxed{5,6,7,8,9}, \boxed{10,11,12}, \boxed{13} & = & I_1, I_2, I_3, I_4 \\ \pi^{(2)}: & \boxed{9,8,7,6,5}, \boxed{10,11,12}, \boxed{1,2,3,4}, \boxed{13} & = & I_2^r, I_3, I_1, I_4 \\ \pi^{(3)}: & \boxed{4,3,2,1}, \boxed{5,6,7,8,9}, \boxed{10,11,12}, \boxed{13} & = & I_1^r, I_2, I_3, I_4 \end{array}$$

The *ordered type* of this is the composition $\alpha = (4, 5, 3, 1)$, and the *unordered type* is the partition $\lambda = (5, 4, 3, 1)$. The *ordered weight* is $U_4U_5U_3U_1$ (where the U_i's are noncommuting) and the *unordered weight* is $u_5u_4u_3u_1$ (where the u_i's are commuting). The *unsigned strips* of $\vec{\pi}$ are $I_1 = \langle 1, 2, 3, 4\rangle$, $I_2 = \langle 5, 6, 7, 8, 9\rangle$, $I_3 = \langle 10, 11, 12\rangle$, $I_4 = \langle 13\rangle$.

Unsigned compression does not uniquely decompose in the same way as Theorem 2.2; we cannot just replace signed arrangements by unsigned arrangements in the theorem statement. If we compress to an unsigned arrangement (replace I_j or I_j^r by j), $(\langle 1, 2, 3, 4\rangle, \langle 2, 3, 1, 4\rangle, \langle 1, 2, 3, 4\rangle)$, it is compressible in this example since it has a strip $(2, 3)$. If we compress to a signed arrangement (replace I_j by j and I_j^r by $-j$), $(\langle 1, 2, 3, 4\rangle, \langle -2, 3, 1, 4\rangle, \langle -1, 2, 3, 4\rangle)$, it's not a bijection because singletons (such as I_4) are the same when reversed. The analog of Theorem 2.2 for unsigned permutations is more complex:

Theorem 2.4. *There is an injection*

$$\Psi_a \;:\; \mathcal{A}_{n,k}^{(g)} \to \mathcal{B}_{k,k}^{(g)} \times \mathcal{C}_{n,k}$$

from unsigned (n, g)-arrangements $\vec{\pi} = (\pi^{(1)}, \ldots, \pi^{(g)}) \in \mathcal{A}_n^{(g)}$ with k strips, to ordered pairs $(\vec{\tau}, \alpha)$, where

(i) $\vec{\tau} = (\tau^{(1)}, \ldots, \tau^{(g)}) \in \mathcal{B}_k^{(g)}$ *is incompressible;*

(ii) $\alpha \in \mathcal{C}_{n,k}$ *is the ordered type of the unsigned arrangement* $\vec{\pi}$*;*

(iii) When $\alpha_j = 1$*, the sign of* j *is* $+1$ *in each of* $\tau^{(1)}, \ldots, \tau^{(g)}$.

Contrast this to Theorem 2.2 for signed arrangements: both input and output arrangements were signed (here $\vec{\pi}$ is unsigned and $\vec{\tau}$ is signed), and there was no (iii).

Next we will state relationships between the strips in $\vec{\sigma}$ and $|\vec{\sigma}|$, as illustrated by Examples 2.1 and 2.3. To state them, we need to define certain partial orders.

Definition 2.5. Let $n \geq 0$ and $\alpha, \beta \in \mathcal{C}_n$. Then β is a *sequential refinement* of α iff β is obtained by concatenating together compositions of $\alpha_1, \alpha_2, \ldots, \alpha_{\ell(\alpha)}$. Further, $\beta \leq \alpha$ in *sequential refinement order* on $\mathcal{C}_n$ iff β is a sequential refinement of α.

Definition 2.6. Let $n \geq 0$ and $\lambda, \mu \in \mathcal{P}_n$. Then μ is a *refinement* of λ iff μ can be obtained by concatenating together partitions of $\lambda_1, \lambda_2, \ldots$ and sorting the parts into nonincreasing order. Further, $\mu \leq \lambda$ in *refinement order* on $\mathcal{P}_n$ iff μ is a refinement of λ.

Definition 2.7. Let α, β be compositions or partitions of $n > 0$. Then $\alpha > \beta$ in *reverse lexicographic order* iff for some k, $\alpha_i = \beta_i$ when $0 < i < k$ and $\alpha_k > \beta_k$. When $n = 0$, there is just one element in $\mathcal{C}_0$ or $\mathcal{P}_0$, so it is equal to itself.

Sequential refinement on compositions, and refinement on partitions, are partial orders. Reverse lexicographic order is a total order that extends both of these partial orders.

In Examples 2.1 and 2.3, the ordered type of $\vec{\sigma}$ is $\beta = (4, 3, 1, 1, 3, 1)$ and the ordered type of $|\vec{\sigma}|$ is $\alpha = (4, 5, 3, 1)$. β is a sequential refinement of α: $4 = 4$, $5 = 3+1+1$, $3 = 3$, $1 = 1$. With unordered types $\mu = (4, 3, 3, 1, 1, 1)$ of $\vec{\sigma}$ and $\lambda = (5, 4, 3, 1)$ of $|\vec{\sigma}|$, we have that μ is a refinement of λ.

Proposition 2.8. *Let $\vec{\sigma}$ be a signed (n, g)-arrangement.*

(i) Let β be the ordered type of $\vec{\sigma}$ and α be the ordered type of $|\vec{\sigma}|$. Then $\beta \leq \alpha$ in sequential refinement order.

(ii) Let μ be the unordered type of $\vec{\sigma}$ and λ be the unordered type of $|\vec{\sigma}|$. Then $\mu \leq \lambda$ in refinement order.

Proof. Strips in $|\vec{\sigma}|$ arise from concatenating one or more consecutive strips in $\vec{\sigma}$, so consecutive strip lengths in $\vec{\sigma}$ are grouped and added together to give lengths in $|\vec{\sigma}|$. $\square$

In the reverse direction, given an unsigned arrangement $\vec{\pi}$, one of the many signed arrangements $\vec{\sigma}$ with $\vec{\pi} = |\vec{\sigma}|$ is as follows; this one is useful because it preserves the type:

Definition 2.9. Let $\vec{\pi} \in \mathcal{A}_n^{(g)}$. The *canonical signage* of $\vec{\pi}$ is the arrangement obtained by decomposing $\vec{\pi}$ into strips, imposing positive signs on the elements in each forwards strip and each singleton (strip of length 1), and negative signs in each reverse strip.

The canonical signage of Example 2.3 is

$$\begin{aligned}
\sigma^{(1)}: &\quad \langle 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 \rangle \\
\sigma^{(2)}: &\quad \langle -9, -8, -7, -6, -5, 10, 11, 12, 1, 2, 3, 4, 13 \rangle \\
\sigma^{(3)}: &\quad \langle -4, -3, -2, -1, 5, 6, 7, 8, 9, 10, 11, 12, 13 \rangle
\end{aligned}$$

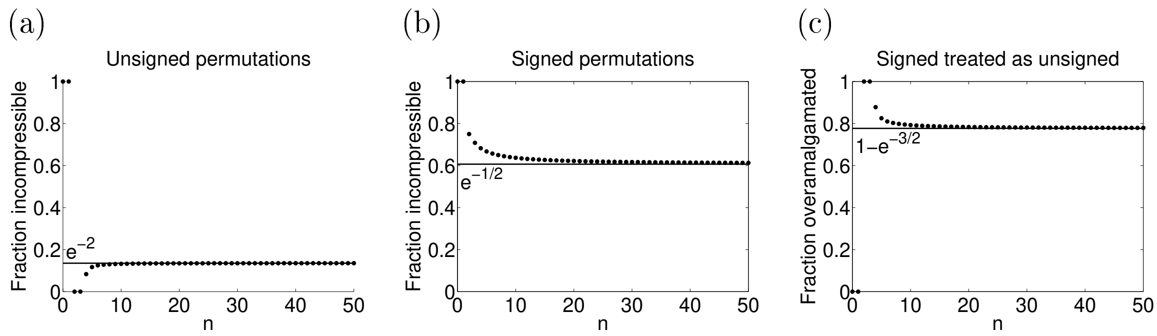

Figure 2: The fraction of arrangements that are incompressible with $g = 2$ genomes of size n, as n increases. (a) Unsigned genomes: the fraction $a^{(2)}_{n,n}/n!$ approaches $\exp(-2) \approx 0.1353$. (b) Signed genomes: the fraction $b^{(2)}_{n,n}/(2^n\, n!)$ approaches $\exp(-\frac{1}{2}) \approx 0.6065$. (c) The fraction of incompressible signed permutations σ that are compressible as unsigned permutations $|\sigma|$ is $1 - 2^n a^{(2)}_{n,n}/b^{(2)}_{n,n}$, which approaches $1 - \exp(-3/2) \approx 0.7769$.

Note that the sign of 13 in $\sigma^{(2)}$ is different than in Example 2.1. In converting unsigned gene orders to signed gene orders, one would typically compute the canonical signage as indicated above, though the true signs of the singletons would remain unclear. See Pevzner and Hannenhalli [13] for additional details. We will discuss it further in Section 8.

3 Strips in signed arrangements

In this section, we derive exact formulas for the number of signed arrangements by ordered type, unordered type, or number of strips, and also asymptotic formulas.

Consider $g \geq 2$ genomes and $n \geq 0$ genes.

Let $B^{(g)}_{\beta}$ denote the number of signed (n, g)-arrangements of ordered type $\beta \in \mathcal{C}_n$ and $b^{(g)}_{\mu}$ denote the number of signed (n, g)-arrangements of unordered type $\mu \in \mathcal{P}_n$.

Note: The notation $b^{(g)}_{\mu}$ is distinguished from $b^{(g)}_{n,k}$ because μ is a partition. So $b^{(g)}_{5,3}$ is the number of length 5 arrangements with 3 strips, while $b^{(g)}_{(5,3)}$ is the number of length 8 arrangements with one length 5 strip and one length 3 strip.

Theorem 3.1. *(i) $b^{(g)}_{0,0} = 1$, and for $k > 0$, we have*

$$b^{(g)}_{k,k} = \sum_{r=1}^{k} (-1)^{k-r} \binom{k-1}{r-1} (2^r r!)^{g-1} \; . \tag{1}$$

In the special case $g = 2$ and $k > 0$, this simplifies as follows, using the integer floor function $\lfloor x \rfloor$; also see Fig. 2(b):

$$b^{(2)}_{k,k} = \left\lfloor k!\, 2^k \, \exp(-\tfrac{1}{2}) \right\rfloor + \left\lfloor (k-1)!\, 2^{k-1} \, \exp(-\tfrac{1}{2}) \right\rfloor + 1 \; . \tag{2}$$

(ii) For $n, k \geq 1$, we have

$$b_{n,k}^{(g)} = \binom{n-1}{k-1} \sum_{r=1}^{k} (-1)^{k-r} \binom{k-1}{r-1} (2^r r!)^{g-1} . \tag{3}$$

The $k = 0$ case is $b_{0,0}^{(g)} = 1$ and $b_{n,0}^{(g)} = 0$ for $n > 0$. Also note $b_{n,k}^{(g)} = 0$ for $n < k$.

(iii) For $\beta \in \mathcal{C}_{n,k}$, $B_\beta^{(g)} = b_{k,k}^{(g)}$.

(iv) For $\mu \in \mathcal{P}_{n,k}$, $b_\mu^{(g)} = b_{k,k}^{(g)} M(\mu) = b_{k,k}^{(g)} \binom{k}{m_1(\mu),\ldots,m_n(\mu)}$.

Theorem 3.2. *Fix $q \geq 0$. As $n \to \infty$, the number of (n, g)-arrangements with exactly q adjacencies, $b_{n,n-q}^{(g)}$, has the following asymptotic form. For $g \geq 2$,*

$$\lim_{n\to\infty} \frac{b_{n,n-q}^{(g)}}{(2^{n-q}(n-q)!)^{g-1}\binom{n-1}{n-q-1}} = \begin{cases} \exp(-\frac{1}{2}) & \text{if } g = 2; \\ 1 & \text{if } g > 2, \end{cases} \tag{4}$$

and for $g = 2$,

$$\lim_{n\to\infty} \frac{b_{n,n-q}^{(2)}}{2^n n!} = \frac{\exp(-\frac{1}{2})}{2^q q!} . \tag{5}$$

The proof of Theorem 3.2 is deferred to Appendix A.1.

Proof of Theorem 3.1. (i) Let $n, k \geq 1$. From the bijection in Theorem 2.2, the number of signed (n, g)-arrangements with k strips is

$$b_{n,k}^{(g)} = |\mathcal{B}_{n,k}^{(g)}| = |\mathcal{B}_{k,k}^{(g)}|\, |\mathcal{C}_{n,k}| = b_{k,k}^{(g)} \binom{n-1}{k-1} . \tag{6}$$

The total number of signed (n, g)-arrangements is $(2^n n!)^{g-1}$, so on summing (6) over all possible numbers of strips ($k = 1$ to n), we obtain, for all $n \geq 1$,

$$(2^n n!)^{g-1} = \sum_{k=1}^{n} b_{n,k}^{(g)} = \sum_{k=1}^{n} b_{k,k}^{(g)} \binom{n-1}{k-1} . \tag{7}$$

This system of equations (7) (one equation for each $n = 1, 2, \ldots$) may be solved for $b_{k,k}^{(g)}$ for $k \geq 1$, giving unique solution (1). This solution is obtained using the first of Riordan's famous inverse relations [27, p. 485, Eq. (1b)]. The proof of (2) is postponed.

(ii) Combining (6) and (1) gives (3). For $n = 0$, the only arrangement is the null arrangement, with 0 strips, giving $b_{0,0}^{(g)} = 1$ and $b_{0,k}^{(g)} = 0$ for $k > 0$.

(iii) Theorem 2.2 gives that the (n, g)-arrangements of ordered type β are in bijection with $\mathcal{B}_{k,k}^{(g)}$, where $k = \ell(\beta)$. So $B_\beta^{(g)} = b_{k,k}^{(g)}$.

(iv) For $\mu \in \mathcal{P}_{n,k}$, the (n, g)-arrangements of unordered type μ come from (n, g)-arrangements of ordered type β where $\beta \in \mathcal{C}_{n,k}$ runs over permutations of the parts of μ. There are $M(\mu) = \binom{k}{m_1(\mu),m_2(\mu),\ldots,m_n(\mu)}$ such values of β, each with $B_\beta^{(g)} = b_{k,k}^{(g)}$. □

To prove (2), we require the following lemma.

Lemma 3.3. *Let* $\exp_k(x) = \sum_{m=0}^{k} x^m/m!$*, where* $k \geq 0$ *is an integer. Then for any integer* $n > 0$*,*

$$\exp_k(-\tfrac{1}{n}) = \frac{\lfloor n^k\, k!\, \exp(-\frac{1}{n}) \rfloor + \frac{1}{2}(1+(-1)^k)}{n^k\, k!} \,. \tag{8}$$

Proof. $\exp_k(-\frac{1}{n}) = \sum_{m=0}^{k}(-\frac{1}{n})^m/m!$ is a partial sum of the Maclaurin series expansion of $\exp(-\frac{1}{n})$. All denominators $m!$ $(0 \leq m \leq k)$ divide $n^k k!$, so $n^k k! \exp_k(-\frac{1}{n})$ is an integer and $n^k k! \exp(-\frac{1}{n}) = n^k k! \exp_k(-\frac{1}{n}) + \epsilon$ where

$$\epsilon = n^k k! \big(\exp(-\tfrac{1}{n}) - \exp_k(-\tfrac{1}{n})\big) = \sum_{m=k+1}^{\infty} \frac{(-1)^m n^{k-m} k!}{m!} \,.$$

ϵ is an alternating series whose first term $(-1)^{k+1}/(k+1)$ has absolute value < 1 (or $= 1$ if $k = 0$). The ratio of term $m+1$ over term m is $-1/(n(m+1))$, with absolute value below 1. So $0 < |\epsilon| < 1$. The sign of ϵ is given by its first term: negative if k is even, positive if k is odd. So the integer $n^k k! \exp_k(-\frac{1}{n})$ may be expressed as

$$n^k\, k!\, \exp_k(-\tfrac{1}{n}) = n^k\, k!\, \exp(-\tfrac{1}{n}) - \epsilon = \lfloor n^k\, k!\, \exp(-\tfrac{1}{n}) \rfloor + \delta \tag{9}$$

where $\delta = \frac{1}{2}(1+(-1)^k) = 1$ if k is even, 0 if k is odd. Dividing (9) by $n^k k!$ gives (8). $\square$

Proof of (2). Plug $g = 2$ and $m = k - r$ into (1), simplify factorials, and apply (8):

$$\begin{aligned}
b^{(2)}_{k,k} &= \sum_{r=1}^{k} (-1)^{k-r} \binom{k-1}{r-1} 2^r r! \;=\; \sum_{m=0}^{k-1} (-1)^m \binom{k-1}{k-m-1} 2^{k-m}(k-m)! \\
&= (k-1)! \sum_{m=0}^{k-1} \frac{(-1)^m 2^{k-m}(k-m)!}{(k-m-1)!\, m!} \;=\; 2^k (k-1)! \sum_{m=0}^{k-1} \frac{(-\frac{1}{2})^m (k-m)}{m!}
\end{aligned}$$

Extend the summation to $m = k$; the $m = k$ term is 0 due to the factor of $k - m$:

$$\begin{aligned}
&= 2^k (k-1)! \sum_{m=0}^{k} \frac{(-\frac{1}{2})^m (k-m)}{m!} \;=\; 2^k k! \sum_{m=0}^{k} \frac{(-\frac{1}{2})^m}{m!} - 2^k (k-1)! \sum_{m=1}^{k} \frac{(-\frac{1}{2})^m}{(m-1)!} \\
&= 2^k k! \exp_k(-\tfrac{1}{2}) + 2^{k-1}(k-1)! \exp_{k-1}(-\tfrac{1}{2}) \\
&= \lfloor 2^k\, k!\, \exp(-\tfrac{1}{2}) \rfloor + \lfloor 2^{k-1}\, (k-1)!\, \exp(-\tfrac{1}{2}) \rfloor + \tfrac{1}{2}\big(2 + (-1)^k + (-1)^{k-1}\big) \,.
\end{aligned}$$

$\square$

4 Generating functions for signed arrangements

In this section, we will define and compute generating functions for the number of signed (n, g)-arrangements by ordered type and by number of strips.

Let $\vec{V} = (V_1, V_2, \ldots)$ be an infinite sequence of noncommuting indeterminates (that commute with t). For convenience, set $V_0 = 1$. Set $V(t) = \sum_{n=1}^{\infty} t^n V_n$.

For a sequence $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ of nonnegative integers (including partitions, compositions, and sequences with 0's), set $V_\beta = V_{\beta_1} V_{\beta_2} \cdots V_{\beta_k}$.

Let $\vec{\sigma} \in \mathcal{B}^{(g)}$ have ordered type β. The *ordered weight* of $\vec{\sigma}$ is $\omega_B(\vec{\sigma}) = V_\beta$. The *ordered weight* of a set of arrangements $S \subset \mathcal{B}^{(g)}$ is $\omega_B(S) = \sum_{\vec{\sigma} \in S} \omega_B(\vec{\sigma})$ and the *graded ordered weight* is $\omega_B(S;t) = \sum_{\vec{\sigma} \in S} t^{n(\vec{\sigma})} \omega_B(\vec{\sigma})$ where if $\vec{\sigma} \in \mathcal{B}_n^{(g)}$ then $n(\vec{\sigma}) = n$.

We define generating functions for the number of arrangements by ordered type:

$$B_n^{(g)}(\vec{V}) = B_n^{(g)}(V_1, V_2, \ldots) = \omega_B(\mathcal{B}_n^{(g)}) = \sum_{\beta \in \mathcal{C}_n} B_\beta^{(g)} V_{\beta_1} V_{\beta_2} \ldots V_{\beta_{\ell(\beta)}} \tag{10}$$

$$B^{(g)}(\vec{V};t) = B^{(g)}(V_1, V_2, \ldots; t) = \omega_B(\mathcal{B}^{(g)}; t) = \sum_{n=0}^{\infty} t^n \sum_{\beta \in \mathcal{C}_n} B_\beta^{(g)} V_{\beta_1} V_{\beta_2} \ldots V_{\beta_{\ell(\beta)}}\,. \tag{11}$$

Eq. (11) is a formal power series in t and an infinite number of noncommuting indeterminates $V_1, V_2, \ldots$. Further, the coefficient of each power of t is a polynomial in $V_1, V_2, \ldots$. Thus, we may work in the noncommutative ring $\mathbb{Z}\langle V_1, V_2, \ldots \rangle[[t]]$. Our main result is

Theorem 4.1. *A generating function to count signed arrangements by ordered type is*

$$B^{(g)}(\vec{V};t) = \sum_{r=0}^{\infty} (2^r r!)^{g-1} \left(\tfrac{V(t)}{1+V(t)} \right)^r \,. \tag{12}$$

Lemma 4.2. $(1+V(t))^{-1}$ *and* $\frac{V(t)}{1+V(t)}$ *are well-defined formal power series, and* $\frac{V(t)}{1+V(t)}$ *is divisible by* t.

Proof. When $V(t)^k$ is expanded as a power series in t, the coefficient of t^n is 0 if $n < k$, and is a polynomial in $V_1, \ldots, V_n$ if $n \geq k$.

$(1+V(t))^{-1} = \sum_{k=0}^{\infty} (-1)^k V(t)^k$ is the formal multiplicative inverse of $1+V(t)$, provided it is well-defined. Indeed, when it is expanded as a power series in t, the coefficient of t^n has contributions only from terms $k = 0, \ldots, n$, so again it is a polynomial in $V_1, \ldots, V_n$.

Finally, $\frac{V(t)}{1+V(t)} = \sum_{k=1}^{\infty} (-1)^{k-1} V(t)^k$ is divisible by t since $V(t)$ is divisible by t. $\square$

Proof of Theorem 4.1. Note that (12) is a well-defined formal power series, even though it is not convergent as an analytic power series. When it is expanded as power series in t, the coefficient of t^n has a finite number of contributions, all from terms $r = 0, 1, \ldots, n$.

Evaluate (11), using the formulas for $B^{(g)}_\beta$ and $b^{(g)}_{k,k}$ from Theorem 3.1:

$$\begin{aligned}
B^{(g)}(\vec{V};t) &= \sum_{n=0}^{\infty} t^n \sum_{k=0}^{n} \sum_{\beta\in\mathcal{C}_{n,k}} B^{(g)}_\beta V_{\beta_1} V_{\beta_2}\cdots V_{\beta_k} \;=\; \sum_{n=0}^{\infty} t^n \sum_{k=0}^{n} \sum_{\beta\in\mathcal{C}_{n,k}} b^{(g)}_{k,k} V_{\beta_1} V_{\beta_2}\cdots V_{\beta_k}\\
&= \sum_{k=0}^{\infty} b^{(g)}_{k,k} \sum_{n=0}^{\infty} \sum_{\beta\in\mathcal{C}_{n,k}} (t^{\beta_1}V_{\beta_1})(t^{\beta_2}V_{\beta_2})\cdots(t^{\beta_k}V_{\beta_k}) \;=\; \sum_{k=0}^{\infty} b^{(g)}_{k,k} V(t)^k\\
&= 1+\sum_{k=1}^{\infty} V(t)^k \sum_{r=1}^{k} (-1)^{k-r}\binom{k-1}{r-1}(2^r r!)^{g-1}\\
&= 1+\sum_{r=1}^{\infty} (2^r r!)^{g-1} \sum_{k=r}^{\infty} (-1)^{k-r}\binom{k-1}{r-1} V(t)^k \;=\; \sum_{r=0}^{\infty}(2^r r!)^{g-1}\left(\frac{V(t)}{1+V(t)}\right)^r . \square
\end{aligned}$$

We consider three specializations of the formal power series (12). These could be computed from Theorem 3.1, but we will show how to do them with specializations since this will be the required method for unsigned arrangements.

1. A generating function to count signed arrangements by unordered types is obtained by allowing $V_1, V_2, \ldots$ to commute. This will be done in detail in Section 6.

2. A generating function to count signed arrangements by size and number of strips. Specializing $V_n \to z$ for $n > 0$ gives $V(t) \to zt/(1-t)$; applying this to (12) gives

$$b^{(g)}(t,z) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} b^{(g)}_{n,k} t^n z^k = \sum_{n=0}^{\infty} b^{(g)}_n(z) t^n = \sum_{r=0}^{\infty} (2^r r!)^{g-1}\left(\frac{zt}{1-t(1-z)}\right)^r . \qquad (13)$$

   Expanding this as a Maclaurin series in t, the coefficients $b_n(z)$ of t^n are

$$b^{(g)}_0(z) = 1 \qquad b^{(g)}_n(z) = \sum_{r=1}^{n} (2^r r!)^{g-1} \tbinom{n-1}{r-1} z^r (1-z)^{n-r} \quad (\text{for } n \geq 1). \qquad (14)$$

3. A generating function, $\mathrm{IB}^{(g)}(t)$, to count incompressible signed arrangements by size. Specialize $V_1 \to 1$ and $V_n \to 0$ for $n > 1$. This gives $V(t) \to t$. Apply it to (12):

$$\mathrm{IB}^{(g)}(t) = \sum_{n=0}^{\infty} b^{(g)}_{n,n} t^n = \sum_{r=0}^{\infty} (2^r r!)^{g-1} \frac{t^r}{(1+t)^r} . \qquad (15)$$

As a side note, the exponential generating function corresponding to this has a nice form when $g = 2$. Let $\mathrm{EIB}(t) = \sum_{n=0}^{\infty} b^{(2)}_{n,n}\, t^n/n!$ be the exponential generating function for the number of incompressible signed $(n,2)$-arrangements.

Theorem 4.3. $\mathrm{EIB}(t) = 1 + \int_0^t \frac{2\exp(-u)}{(1-2u)^2}\, du.$

Proof.

$$\mathrm{EIB}'(t)\exp(t) = \sum_{k=1}^{\infty} b^{(2)}_{k,k}\frac{t^{k-1}}{(k-1)!}\sum_{m=1}^{\infty}\frac{t^{m-1}}{(m-1)!} = \sum_{n=1}^{\infty}\frac{t^{n-1}}{(n-1)!}\sum_{k=1}^{n}\binom{n-1}{k-1}b^{(2)}_{k,k}$$

where we collected by powers t^{n-1}, with $(n-1) = (k-1) + (m-1)$. Next plug in (7):

$$= \sum_{n=1}^{\infty} 2^n n!\frac{t^{n-1}}{(n-1)!} = \sum_{n=1}^{\infty} 2^n n t^{n-1} = \frac{2}{(1-2t)^2}$$

$$\mathrm{EIB}'(t) = \sum_{k=1}^{\infty} b^{(2)}_{k,k}\frac{t^{k-1}}{(k-1)!} = \frac{2\exp(-t)}{(1-2t)^2}$$

Integrate and use initial condition $\mathrm{EIB}(0) = b^{(2)}_{0,0} = 1$ to obtain $\mathrm{EIB}(t)$ as stated. □

5 Strips in unsigned arrangements

In this section, we will obtain a generating function for enumeration of unsigned arrangements by ordered type. We will use this to determine formulas for the number of unsigned arrangements by type, or with a specified number of strips. The computations are considerably more complicated than for signed arrangements. Section 5.1 gives the notation for the unsigned case and develops a map between the weight of an unsigned arrangements and all signed arrangements arising from implanting signs in it. Section 5.2 gives generating functions for unsigned arrangements by ordered type and by number of strips.

5.1 Weights on adding signs to unsigned arrangements

We adopt notation similar to that of Section 4. Essentially, symbols $\mathcal{B}$, b, β, V, μ, for signed arrangements will be replaced by $\mathcal{A}$, a, α, U, λ, for unsigned arrangements, including font, capitalization, and sub/superscript variations.

Let $\vec{U} = (U_1, U_2, \ldots)$ be an infinite sequence of noncommuting indeterminates (that commute with t). For convenience, set $U_0 = 1$. Set $U(t) = \sum_{n=1}^{\infty} t^n U_n$.

Let $\vec{\pi} \in \mathcal{A}^{(g)}$ with ordered type α. The *ordered weight* of $\vec{\pi}$ is $\omega_A(\vec{\pi}) = U_\alpha = U_{\alpha_1}U_{\alpha_2}\cdots$.

The *ordered weight* of a set $S \subset \mathcal{A}^{(g)}$ of arrangements is $\omega_A(S) = \sum_{\vec{\pi}\in S}\omega_A(\vec{\pi})$ and the *graded ordered weight* is $\omega_A(S;t) = \sum_{\vec{\pi}\in S} t^{n(\vec{\pi})}\omega_A(\vec{\pi})$.

The generating functions for counting unsigned arrangements by ordered type are

$$A^{(g)}_n(\vec{U}) = A^{(g)}_n(U_1, U_2, \ldots) = \omega_A(\mathcal{A}^{(g)}_n) = \sum_{\alpha\in\mathcal{C}_n} A^{(g)}_\alpha U_{\alpha_1}U_{\alpha_2}\ldots U_{\alpha_{\ell(\alpha)}} \tag{16}$$

$$A^{(g)}(\vec{U};t) = A^{(g)}(U_1, U_2, \ldots; t) = \omega_A(\mathcal{A}^{(g)};t) = \sum_{n=0}^{\infty} t^n \sum_{\alpha\in\mathcal{C}_n} A^{(g)}_\alpha U_{\alpha_1}U_{\alpha_2}\ldots U_{\alpha_{\ell(\alpha)}}\ . \tag{17}$$

This is a formal power series in an infinite number of noncommuting indeterminates, in the ring $\mathbb{Z}\langle U_1, U_2, \ldots\rangle[[t]]$. In Section 5.2, we will derive a formula for this series and apply it to get an explicit formula for $a_{n,k}^{(g)}$, the number of (n,g)-arrangements with k strips, as well as generating functions for it and an asymptotic formula. But first, in this section, we develop the machinery to relate the weight of an unsigned arrangement to the weight of all signed arrangements that arise by implanting signs into it.

Implanting signs in an unsigned strip that is forwards in all genomes.

The (n,g)-*identity arrangement* is $\mathrm{id}_n^{(g)} = (\mathrm{id}_n, \ldots, \mathrm{id}_n)$ (g copies of $\langle 1, \ldots, n\rangle$).

Consider an unsigned strip of length n, w.l.o.g. $\mathrm{id}_n^{(g)}$. Signs may be implanted to form a signed (n,g)-arrangement $\vec{\sigma} = (\sigma^{(1)}, \ldots, \sigma^{(g)})$: $\sigma^{(i)} = (\epsilon_{i1}1, \epsilon_{i2}2, \ldots, \epsilon_{in}n)$ for $i = 1, \ldots, g$, where $\epsilon_{1,j} = 1$ and each $\epsilon_{ij} \in \{+1, -1\}$ for $i = 2, \ldots, g$.

The sign vector of j is $\vec{\epsilon}_j = (\epsilon_{1j}, \ldots, \epsilon_{gj})$. Each entry $j = 1, \ldots, n$ has 2^{g-1} possible sign combinations. Let $\vec{\epsilon}_+ = (+1, \ldots, +1)$ (of length g) consist of all positive signs. Set

$$G = 2^{g-1} - 1\,, \qquad \widetilde{G} = 2^{1-g} - 1\,. \tag{18}$$

There are G possible sign vectors besides $\vec{\epsilon}_+$. For later use, we note that

$$G = -\widetilde{G}/(\widetilde{G}+1)\,, \qquad \widetilde{G} = -G/(G+1)\,, \qquad \widetilde{G}+1 = \frac{1}{G+1}\,. \tag{19}$$

A run of m consecutive entries with sign vector $\vec{\epsilon}_+$ forms a signed strip of length m. Each entry with sign vector different from $\vec{\epsilon}_+$ forms a signed strip on one element.

Let $0 < j_1 < \cdots < j_k = n+1$ where $j_1, \ldots, j_{k-1}$ are the positions for which $\vec{\epsilon}_j \neq \vec{\epsilon}_+$. Let $\beta = (j_1, j_2 - j_1, j_3 - j_2, \ldots, j_k - j_{k-1})$. Then as a signed permutation, we form strips of lengths $(\beta_1 - 1, 1, \beta_2 - 1, 1, \ldots, \beta_{k-1} - 1, 1, \beta_k - 1)$ (except that we omit any 0's that arise from $\beta_r - 1$ with $\beta_r = 1$).

Example 5.1. Consider adding signs to an unsigned strip of length $n = 9$ in 3 genomes:

$$\begin{array}{rl}
\sigma^{(1)}: & \boxed{1,2\,\boxed{3}},\ \boxed{4},\ \boxed{5,6,7\,\boxed{8}},\ \boxed{9} \\
\sigma^{(2)}: & \boxed{1,2\,\boxed{-3}},\ \boxed{-4},\ \boxed{5,6,7\,\boxed{8}},\ \boxed{9} \\
\sigma^{(3)}: & \boxed{1,2\,\boxed{3}},\ \boxed{4},\ \boxed{5,6,7\,\boxed{-8}},\ \boxed{9}
\end{array}$$

The positions that are not all positive (sign $\neq \vec{\epsilon}_+$) are $3, 4, 8$ (shown with bold boxes around them), and we add on $n + 1 = 10$ to this list (though it is not part of the permutations). The successive differences of these positions give a composition of $n+1$: $\beta = (3, 4-3, 8-4, 10-8) = (3, 1, 4, 2)$. Note that the arrangement alternates between positive strips (possibly of length 0) and non-positive positions, and each part of the composition represents joining a strip with the non-positive position terminating it.

The strip lengths are $(3-1, 1, 1-1, 1, 4-1, 1, 2-1) = (2, 1, 0, 1, 3, 1, 1)$, and we omit all zeros to obtain $(2, 1, 1, 3, 1, 1)$. The ordered weight of this is $V_2V_1V_0V_1V_3V_1V_1 = V_2V_1V_1V_3V_1V_1$, while the unordered weight is $v_3v_2{v_1}^4$.

If all entries but 3, 4, 8 have sign vector $\vec{\epsilon}_+$, then for entries 3, 4, and 8, we may independently choose any of $G = 2^2 - 1 = 3$ sign vectors not equal to $\vec{\epsilon}_+$, and get the same partition into strips as shown above (but with different sign vectors on entries 3, 4, 8). So there are $G^3 = 27$ signages obtained from signs $\neq \vec{\epsilon}_+$ in precisely those positions.

Implanting signs in an unsigned strip that is backwards in some genomes.

Consider any unsigned strip of length $n > 1$ in g genomes. The canonical sign vector $\vec{\epsilon}_c = (\epsilon_1, \ldots, \epsilon_g)$ has $\epsilon_i = +1$ if the strip is forwards in genome i and $\epsilon_i = -1$ if it's backwards. The canonical signage assigns sign ϵ_i to all entries in that strip in genome i. The weights and counts of all signages where sign vectors $\neq \vec{\epsilon}_c$ are implanted at certain entries is the same as computed above for implanting signs $\neq \vec{\epsilon}_+$ at those entries in $\mathrm{id}_n^{(g)}$. In Example 5.1, if the strip is backwards in the third genome, the canonical signage is

$$\begin{array}{ll} \sigma^{(1)}: & 1,2,3,4,5,6,7,8,9 \\ \sigma^{(2)}: & 1,2,3,4,5,6,7,8,9 \\ \sigma^{(3)}: & -9,-8,-7,-6,-5,-4,-3,-2,-1 \end{array}$$

and the sign modifications on entries $3, 4, 8$ corresponding to the ones in Example 5.1 are

$$\begin{array}{ll} \sigma^{(1)}: & \boxed{1,2}\boxed{3}\,,\ \boxed{4}\,,\ \boxed{5,6,7}\boxed{8}\,,\ \boxed{9} \\ \sigma^{(2)}: & \boxed{1,2}\boxed{\text{-}3}\,,\ \boxed{\text{-}4}\,,\ \boxed{5,6,7}\boxed{8}\,,\ \boxed{9} \\ \sigma^{(3)}: & \boxed{\text{-}9}\,,\ \boxed{8}\boxed{\text{-}7,\text{-}6,\text{-}5}\,,\ \boxed{\text{-}4}\,,\ \boxed{\text{-}3}\boxed{\text{-}2,\text{-}1} \end{array}$$

Theorem 5.2. *(i) The ordered weight of all signages of unsigned* $\mathrm{id}_n^{(g)}$ *($n > 0$) is*

$$\begin{aligned} \phi(U_n) &= \sum_{k=1}^{n} \sum_{\beta \in \mathcal{C}_{n+1,k}} V_{\beta_1 - 1} \cdot GV_1 \cdot V_{\beta_2 - 1} \cdot GV_1 \cdots V_{\beta_k - 1} \\ &= \sum_{k=1}^{n} G^{k-1} \sum_{\beta \in \mathcal{C}_{n+1,k}} V_{\beta_1 - 1, 1, \beta_2 - 1, 1, \cdots, \beta_k - 1} \ . \end{aligned} \tag{20}$$

(ii) Let $\vec{\pi}$ be an unsigned (n, g)-arrangement with ordered type α. The ordered weight of all signages of $\vec{\pi}$ is $\phi(U_{\alpha_1})\phi(U_{\alpha_2})\cdots$.

Proof. Part (i) is clear from the example above. There are $k-1$ entries with non-canonical sign: $\beta_1, \beta_1 + \beta_2, \ldots, \beta_1 + \cdots + \beta_{k-1}$. Entry $\beta_1 + \cdots + \beta_k = n + 1$ terminates the strip.

For part (ii), the signages subdivide the original strips of $\vec{\pi}$. We choose one of the signages of the first strip (as ordered in $\pi^{(1)}$, one of the signages of the second strip, and so on, independently for each strip. The ordered type of the signage is the concatenation of the ordered types of the signages applied to each original strip, while the unordered type is obtained from this by sorting the parts. So we apply part (i) to each separate strip of $\vec{\pi}$ (relabelling the elements from $1, 2, \ldots$ into those of the strip) and combine the weights of the strips together by noncommutative multiplication of their signed weights in the same order as the strips are in $\pi^{(1)}$. ∎

Define a ring homomorphism $\phi : \mathbb{Q}\langle U_1, U_2, \ldots\rangle \to \mathbb{Q}\langle V_1, V_2, \ldots\rangle$ by defining $\phi(U_i)$ via (20). U_i's are generators, so this extends to the whole ring via $\phi(f+h) = \phi(f) + \phi(h)$ and $\phi(fh) = \phi(f)\phi(h)$. We shall see that this is actually a ring isomorphism. It induces a homomorphism $\phi : \mathbb{Q}\langle U_1, U_2, \ldots\rangle[[t]] \to \mathbb{Q}\langle V_1, V_2, \ldots\rangle[[t]]$ by applying ϕ to the coefficient of each power of t.

Corollary 5.3. $\phi\big(A^{(g)}(\vec{U};t)\big) = B^{(g)}(\vec{V};t)$.

We will develop additional formulas for ϕ and its inverse, so that we may compute generating functions for signed arrangements in terms of the generating functions for unsigned arrangements. A recursion for $\phi(U_n)$ is easy to obtain from (20):

Theorem 5.4. *For $n \geq 1$,*

$$\phi(U_n) = V_n + \sum_{r=0}^{n-1} V_r \cdot GV_1 \cdot \phi(U_{n-1-r}) = V_n + \sum_{r=0}^{n-1} \phi(U_{n-1-r}) \cdot GV_1 \cdot V_r\ . \tag{21}$$

Proof. The k sum in (20) has one term for $k = 1$, namely V_n (corresponding to $\beta = (n+1)$). For $k > 1$, we factor off $V_r \cdot GV_1$ from the left (where $r = \beta_1 - 1 \geq 0$) or $GV_1 \cdot V_r$ from the right (where $r = \beta_k - 1 \geq 0$) to obtain a sum of the exact same form with a smaller n (namely $n-r-1$). For the terms where some $\beta_i = 1$, note that $\phi(U_0) = \phi(1) = 1 = V_0$. □

For $\alpha \in \mathcal{C}_n$, let $U_\alpha = U_{\alpha_1} U_{\alpha_2} \cdots$. Then

$$\phi(U_\alpha) = \phi(U_{\alpha_1})\phi(U_{\alpha_2}) \cdots = \sum_{\beta \in \mathcal{C}_n} H_{\alpha\beta}(G) V_\beta \tag{22}$$

where we plug in (20), expand the products, collect terms, and obtain transition matrix $H(G)$ from the coefficients. For $n > 0$, $H(G)$ is a $2^{n-1} \times 2^{n-1}$ matrix, indexed by compositions $\alpha, \beta \in \mathcal{C}_n$. (For $n = 0$, it is 1×1.) We list the row α and column β indices in reverse lexicographic order on $\mathcal{C}_n$ (we will see below that any extension of sequential refinement order is suitable); see Definitions 2.5 and 2.7. Each matrix entry $H_{\alpha\beta}(G)$ is a polynomial in G with nonnegative integer coefficients. If $\vec{\pi}$ is an unsigned (n, g)-arrangement of ordered type α, then $H_{\alpha\beta}(G)$ gives the number of signages of $\vec{\pi}$ with ordered type β.

Next we develop formulas to compute ϕ^{-1}. Recall that we defined generating functions $U(t) = \sum_{n=1}^{\infty} t^n U_n$ and $V(t) = \sum_{n=1}^{\infty} t^n V_n$. Note that $U_0 = V_0 = 1$ are not included in $U(t), V(t)$, so we use $1 + U(t)$ or $1 + V(t)$ to include the constant term when necessary.

Theorem 5.5. *(i) ϕ is invertible, hence it is a ring isomorphism.*

(ii) In sequential refinement order on compositions, $H(G)$ is lower triangular with 1's on the diagonal.

(iii) $H(G)^{-1} = H(\widetilde{G})$, where $\widetilde{G} = -G/(G+1) = 2^{-(g-1)} - 1$. Thus for $\alpha \in \mathcal{C}_n$,

$$\phi^{-1}(V_\alpha) = \phi^{-1}(V_{\alpha_1})\phi^{-1}(V_{\alpha_2}) \cdots = \sum_{\beta \in \mathcal{C}_n} H_{\alpha\beta}(\widetilde{G}) U_\beta \tag{23}$$

(iv) A practical way to compute $\phi^{-1}(V_\alpha)$ is via the product in (23) and the recursion, for $n \geq 1$,

$$\phi^{-1}(V_n) = U_n + \sum_{r=0}^{n-1} U_r \cdot \widetilde{G}U_1 \cdot \phi^{-1}(V_{n-1-r}) = U_n + \sum_{r=0}^{n-1} \phi^{-1}(V_{n-1-r}) \cdot \widetilde{G}U_1 \cdot U_r \quad (24)$$

(v) Recursions (21) and (24) have solutions in terms of generating functions:

$$\begin{aligned}\phi(U(t)) &= \Big(1 - \big(1+V(t)\big)GV_1t\Big)^{-1}\Big(\big(1+V(t)\big)GV_1t + V(t)\Big) \\ &= \Big(GV_1t\big(1+V(t)\big) + V(t)\Big)\Big(1 - GV_1t\big(1+V(t)\big)\Big)^{-1} \end{aligned} \quad (25)$$

$$\begin{aligned}\phi^{-1}(V(t)) &= \Big(1 - \big(1+U(t)\big)\widetilde{G}U_1t\Big)^{-1}\Big(\big(1+U(t)\big)\widetilde{G}U_1t + U(t)\Big) \\ &= \Big(\widetilde{G}U_1t\big(1+U(t)\big) + U(t)\Big)\Big(1 - \widetilde{G}U_1t\big(1+U(t)\big)\Big)^{-1} \end{aligned} \quad (26)$$

(vi) **Duality:** *Let $f(z; x_1, x_2, \ldots; y_1, y_2, \ldots) \in \mathbb{Q}(z)\langle x_1, x_2, \ldots; y_1, y_2, \ldots\rangle$.*
Then $\quad f\big(G; \phi(U_1), \phi(U_2), \ldots; V_1, V_2, \ldots\big) = 0 \quad$ *in* $\quad \mathbb{Q}(G)\langle V_1, V_2, \ldots\rangle$
iff $\quad f\big(\widetilde{G}; \phi^{-1}(V_1), \phi^{-1}(V_2), \ldots; U_1, U_2, \ldots\big) = 0 \quad$ *in* $\quad \mathbb{Q}(\widetilde{G})\langle U_1, U_2, \ldots\rangle$.

(Note that duality requires using the formal variables G and $\widetilde{G}$; one may not *plug in specific values of g.)*

Note: Examples of duality include (21) vs. (24); (22) vs. (23); and (25) vs. (26).

Proof. (i,ii) By (21), $\phi(U_\alpha) = \phi(U_{\alpha_1})\phi(U_{\alpha_2})\cdots = V_\alpha + \cdots$ where the remaining terms are a linear combination of V_β's with β less than α in sequential refinement order. So the transition matrix for ϕ in the basis from U_α's to V_β's is triangular with 1's on the diagonal. Thus, ϕ is invertible.

(iii,iv,vi) The recursion (21) may be recast in terms of $U(t), V(t)$ in either of two ways:

$$\begin{aligned}\phi(U(t)) &= V(t) + \big(1+V(t)\big)GV_1t\Big(1+\phi\big(U(t)\big)\Big) \\ &= V(t) + \Big(1+\phi\big(U(t)\big)\Big)GV_1t\big(1+V(t)\big) \,. \end{aligned} \quad (27)$$

Isolating the leading $V(t)$ term in each formula gives

$$\begin{aligned}V(t) &= \phi\big(U(t)\big) - \big(1+V(t)\big)GV_1t\Big(1+\phi\big(U(t)\big)\Big) \\ &= \phi\big(U(t)\big) - \Big(1+\phi\big(U(t)\big)\Big)GV_1t\big(1+V(t)\big) \,. \end{aligned}$$

Apply ϕ^{-1}. Note that ϕ is multiplicative and invertible so ϕ^{-1} is too, and $\phi^{-1}(V_1) = \frac{U_1}{G+1}$:

$$\begin{aligned}\phi^{-1}(V(t)) &= U(t) - \big(1+U(t)\big)\frac{G}{G+1}U_1t\Big(1+\phi^{-1}\big(V(t)\big)\Big) \\ &= U(t) - \Big(1+\phi^{-1}\big(V(t)\big)\Big)\frac{G}{G+1}U_1t\big(1+U(t)\big) \,. \end{aligned}$$

Set $\widetilde{G} = -G/(G+1)$ and rewrite that as

$$\begin{aligned}\phi^{-1}(V(t)) &= U(t) + \big(1+U(t)\big)\widetilde{G}U_1 t\Big(1+\phi^{-1}\big(V(t)\big)\Big) \\ &= U(t) + \Big(1+\phi^{-1}\big(V(t)\big)\Big)\widetilde{G}U_1 t\big(1+U(t)\big) . \end{aligned} \tag{28}$$

Expand (28) as a series in t and take the coefficient of t^n to get recursion (24); this proves (iv). Alternately, compare equations (27) and (28). $\phi(U_i)$ $(i \geq 1)$, V_j $(j \geq 1)$, G in the former have been interchanged with $\phi^{-1}(V_i)$, U_j, $\widetilde{G}$ in the latter. In the same manner as recursion (21) leads to an equation (27) in the generating functions, we apply these interchanges to obtain that generating function equation (28) leads to recursion (24).

Evaluating recursion (21) leads to an expansion $\phi(U_n) = \sum_\beta H_{(n),\beta}(G)V_\beta$ of form (22) (with $\alpha = (n)$). Evaluating recursion (24) leads to a similar expansion but with the interchanges above, $\phi^{-1}(V_n) = \sum_\beta H_{(n),\beta}(\widetilde{G})U_\beta$ (Eq. (23) with $\alpha = (n)$).

Then the product $\phi(U_\alpha) = \phi(U_{\alpha_1})\phi(U_\alpha)\cdots$ expanded as a linear combination of V_β's, and $\phi^{-1}(V_\alpha) = \phi^{-1}(V_{\alpha_1})\phi^{-1}(V_{\alpha_2})\cdots$ expanded as a linear combination of U_β's, have similar coefficients except that G in the former coefficients is replaced by $\widetilde{G}$ in the latter. This gives (23), proving (iii). More generally, it leads to a duality theorem (vi).

(v) Eq. (27) can be solved for $\phi(U(t))$, and (28) can be solved for $\phi^{-1}(V(t))$. We show the first equality in (25); the other parts of (25) and (26) are shown similarly. By (27),

$$\begin{aligned}\phi(U(t)) &= V(t) + \big(1+V(t)\big)G\,V_1\,t\Big(1+\phi\big(U(t)\big)\Big) \\ &= V(t) + \big(1+V(t)\big)G\,V_1\,t + \big(1+V(t)\big)G\,V_1\,t\,\phi\big(U(t)\big)\end{aligned}$$

so

$$\Big(1-\big(1+V(t)\big)G\,V_1\,t\Big)\phi\big(U(t)\big) = V(t) + \big(1+V(t)\big)G\,V_1\,t\ ,$$

giving

$$\phi(U(t)) = \Big(1-\big(1+V(t)\big)G\,V_1\,t\Big)^{-1}\Big(V(t)+\big(1+V(t)\big)G\,V_1\,t\Big)\ . \qquad \square$$

In Section 7, we will show how to use the preceding results to compute the number of unsigned (n,g)-arrangements by type.

Our next goal is to compute a formal power series for the graded weights of all unsigned arrangements, but first we need to compute ϕ^{-1} on various expressions.

Lemma 5.6.

$$\phi^{-1}(V_1) = \tfrac{U_1}{G+1} = (\widetilde{G}+1)U_1 \quad \textit{and} \quad \phi^{-1}(GV_1) = -\widetilde{G}U_1 \tag{29}$$

$$\phi^{-1}\big(1+V(t)\big) = (1-\big(1+U(t)\big)\widetilde{G}U_1 t)^{-1}(1+U(t)) \tag{30}$$

$$\phi^{-1}(\big(1+V(t)\big)^{-1}) = \big(1+U(t)\big)^{-1} - \widetilde{G}U_1 t \tag{31}$$

$$\phi^{-1}\left(\frac{V(t)}{1+V(t)}\right) = \frac{U(t)}{1+U(t)} + \widetilde{G}U_1 t. \tag{32}$$

Proof. Eq. (29) is the $n=1$ cases of (21) and (24). They are related using (19).

Note that $\phi^{-1}(1) = 1 = \big(1-(1+U(t))\widetilde{G}U_1 t\big)\big(1-(1+U(t))\widetilde{G}U_1 t\big)^{-1}$. Add this to (26) and simplify the numerator to get (30). Simplify the reciprocal of (30) to get (31).

Subtract both sides of (31) from $\phi^{-1}(1)=1$. Substitute $1-(1+V(t))^{-1} = \frac{V(t)}{1+V(t)}$ and $1-(1+U(t))^{-1} = \frac{U(t)}{1+U(t)}$ to get (32): $\square$

5.2 Generating functions for unsigned arrangements

Theorem 5.7. *A generating function to count unsigned arrangements by ordered type is*

$$A^{(g)}(\vec{U};t) = \sum_{n=0}^{\infty}\sum_{\alpha\in\mathcal{C}_n} A^{(g)}_{\alpha} U_\alpha = \sum_{r=0}^{\infty}(2^r r!)^{g-1}\left(\frac{U(t)}{1+U(t)} + \widetilde{G}U_1 t\right)^r . \qquad (33)$$

Proof.

$$\begin{aligned}
A^{(g)}(\vec{U};t) &= \phi^{-1}\big(B^{(g)}(\vec{V};t)\big) && \text{by Corollary 5.3 and Theorem 5.5}\\
&= \sum_{r=0}^{\infty}(2^r r!)^{g-1}\left(\phi^{-1}\left(\frac{V(t)}{1+V(t)}\right)\right)^r && \text{by Theorem 4.1}\\
&= \sum_{r=0}^{\infty}(2^r r!)^{g-1}\left(\frac{U(t)}{1+U(t)} + \widetilde{G}U_1 t\right)^r && \text{by (32).} \qquad \square
\end{aligned}$$

Now we consider three specializations of this formal power series for $A^{(g)}(\vec{U};t)$.

1. A generating function to count unsigned arrangements by unordered types is obtained by allowing $U_1, U_2, \ldots$ to commute. This will be done in detail in Section 6.

2. A generating function to count incompressible unsigned permutations by size is obtained by specializing (33) with $U_1 \to 1$ and $U_n \to 0$ for $n>1$. This specialization gives $U(t)\to t$ and

$$\frac{U(t)}{1+U(t)} + \widetilde{G}u_1 t \to \frac{t}{1+t} + \widetilde{G}t = \frac{t(1+\widetilde{G}+\widetilde{G}t)}{1+t} = \frac{t(1-Gt)}{(1+G)(1+t)} = \frac{t(1-Gt)}{2^{g-1}(1+t)}$$

where we made use of (18–19). Plugging this into (33) gives the specialization

$$\mathrm{IA}^{(g)}(t) = \sum_{n=0}^{\infty} a^{(g)}_{n,n} t^n = \sum_{r=0}^{\infty}(2^r r!)^{g-1}\left(\frac{t(1-Gt)}{2^{g-1}(1+t)}\right)^r = \sum_{r=0}^{\infty} r!^{g-1}\left(\frac{t(1-Gt)}{1+t}\right)^r . \qquad (34)$$

For the $g=2$ case, the sequence $a^{(2)}_{n,n}$ is listed in the On-Line Encyclopedia of Integer Sequences, A002464 [31]. Further references will follow Theorem 5.8.

3. Specializing $U_n \to z$ for $n>0$ in (33) gives a generating function for the number of unsigned arrangements by size and number of strips:

$$a^{(g)}(t,z) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a^{(g)}_{n,k}\, t^n z^k = \sum_{n=0}^{\infty} a^{(g)}_n(z)\, t^n \qquad \text{where } a^{(g)}_n(z) = \sum_{k=0}^{n} a^{(g)}_{n,k} z^k .$$

Theorem 5.8. *For $g \geq 2$, $n \geq 1$, and $1 \leq k \leq n$,*

$$a^{(g)}(t,z) = \sum_{r=0}^{\infty} r!^{g-1} \left(\frac{zt\big(1+Gt(1-z)\big)}{1-t(1-z)} \right)^r \tag{35}$$

$$a_n^{(g)}(z) = \sum_{r=0}^{n} r!^{g-1} z^r (1-z)^{n-r} \sum_{i=0}^{\min(r,n-r)} G^i \binom{r}{i} \binom{n-i-1}{r-1} \tag{36}$$

$$a_{n,k}^{(g)} = \sum_{r=0}^{k} r!^{g-1} (-1)^{k-r} \binom{n-r}{k-r} \sum_{i=0}^{\min(r,n-r)} G^i \binom{r}{i} \binom{n-i-1}{r-1} \tag{37}$$

Initial conditions are $a_0^{(g)}(z) = 1$, $a_{0,0}^{(g)} = 1$, $a_{0,k}^{(g)} = 0$ for $k > 0$, and $a_{n,0}^{(g)} = 0$ for $n > 0$.

Note: In different notation than ours, Riordan, 1965 [28, p. 710, Eq. (17)] states the $g = 2$ case of (35); he attributes the result to Carlitz. Also in different notation, Abramson and Moser, 1967 [1, p. 1249, Eq. (i)] prove a formula for the $g = 2$ case of (37).

Proof. In (33), specialize $U_n \to z$ for $n > 0$, giving $U(t) \to zt/(1-t)$. Then

$$\frac{U(t)}{1+U(t)} + \widetilde{G}U_1 t \to \frac{zt}{1-t+zt} + \widetilde{G}zt = \frac{zt\big(1+Gt(1-z)\big)}{(G+1)\big(1-t(1-z)\big)} = \frac{zt\big(1+Gt(1-z)\big)}{2^{g-1}\big(1-t(1-z)\big)} .$$

Plugging into (33) and cancelling the powers of 2 gives (35). Expand (35) as a formal power series in t to obtain $a_n^{(g)}(z)$ as the coefficient of t^n. Expand the numerator using the Binomial Theorem, and the denominator using the negative binomial series $(1-y)^{-r} = \sum_{j=0}^{\infty} \binom{r+j-1}{r-1} y^j$, with $y = t(1-z)$:

$$a^{(g)}(t,z) = 1 + \sum_{r=1}^{\infty} r!^{g-1} t^r z^r \left(\sum_{i=0}^{r} \binom{r}{i} (Gt(1-z))^i \right) \left(\sum_{j=0}^{\infty} \binom{r+j-1}{r-1} (t(1-z))^j \right) \tag{38}$$

Collect (38) by powers t^n, where $n = r+i+j$ and $j = n-r-i$:

$$\begin{aligned} a^{(g)}(t,z) &= 1 + \sum_{n=1}^{\infty} t^n \sum_{r=1}^{n} r!^{g-1} z^r \sum_{i=0}^{\min(r,n-r)} \binom{r}{i} (G(1-z))^i \binom{n-i-1}{r-1} (1-z)^{n-r-i} \\ &= 1 + \sum_{n=1}^{\infty} t^n \sum_{r=1}^{n} r!^{g-1} z^r (1-z)^{n-r} \sum_{i=0}^{\min(r,n-r)} G^i \binom{r}{i} \binom{n-i-1}{r-1} \end{aligned}$$

Take the coefficient of t^n to get (36). Finally, expand this as a polynomial in z and compute the coefficient $a_{n,k}^{(g)}$ of z^k to prove (37). Note that the coefficient of z^k in $z^r(1-z)^{n-r}$ equals $(-1)^{k-r}\binom{n-r}{k-r}$ if $0 \leq r \leq k \leq n$ and equals 0 otherwise. ∎

The following theorem states asymptotic formulas for $a_{n,k}^{(g)}$; the proof is postponed to Appendix A.2.

Theorem 5.9. *For $g \geq 2$,*

$$\lim_{n\to\infty} \frac{a^{(g)}_{n,n-q}}{n!(n-q)!^{g-2}2^{q(g-1)}/q!} = \begin{cases} \exp(-2) & \text{if } g = 2; \\ 1 & \text{if } g > 2, \end{cases} \tag{39}$$

and for $g = 2$,

$$\lim_{n\to\infty} \frac{a^{(2)}_{n,n-q}}{n!} = \frac{2^q \exp(-2)}{q!} . \tag{40}$$

Note: Eq. (40) was proved, in different notation than ours, by Wolfowitz, 1944 [34]. Kaplansky, 1945 [15] gave two additional subdominant terms. See Fig. 2(a) for a plot of the $g = 2$, $q = 0$ case.

6 Generating functions by unordered type

In this section, we give generating functions for the number of signed or unsigned arrangements by unordered type. The results of Sections 4-5 for ordered types have analogs for unordered types, obtained by allowing the variables to commute. We will use lowercase variables for the commutative case. Let $\vec{u} = (u_1, u_2, \ldots)$ and $\vec{v} = (v_1, v_2, \ldots)$ be infinite sequences of commuting indeterminates. For convenience, set $u_0 = v_0 = 1$. Set $u(t) = \sum_{n=1}^{\infty} t^n u_n$ and $v(t) = \sum_{n=1}^{\infty} t^n v_n$.

Definition 6.1. The *commutative specialization* of a function in noncommuting variables $U_1, U_2, \ldots$ or $V_1, V_2, \ldots$ is obtained by specializing $U_n \to u_n$ and $V_n \to v_n$ for all $n \geq 1$.

A signed arrangement $\vec{\sigma} \in \mathcal{B}^{(g)}$ with unordered type μ has *unordered weight* $\omega_b(\vec{\sigma}) = v_\mu = v_{\mu_1} v_{\mu_2} \cdots$. An unsigned arrangement $\vec{\pi} \in \mathcal{A}^{(g)}$ of unordered type λ has *unordered weight* $\omega_a(\vec{\pi}) = u_\lambda = u_{\lambda_1} u_{\lambda_2} \cdots$. These are extended to the (graded) unordered weight of sets of arrangements analogously to the ordered case.

The generating functions for counting *signed arrangements* by unordered type are

$$b^{(g)}_n(\vec{v}) = b^{(g)}_n(v_1, v_2, \ldots) = \omega_b(\mathcal{B}^{(g)}_n) = \sum_{\mu \in \mathcal{P}_n} b^{(g)}_\mu v_{\mu_1} v_{\mu_2} \ldots v_{\mu_{\ell(\mu)}} \tag{41}$$

$$b^{(g)}(\vec{v}; t) = b^{(g)}(v_1, v_2, \ldots; t) = \omega_b(\mathcal{B}^{(g)}; t) = \sum_{n=0}^{\infty} t^n \sum_{\mu \in \mathcal{P}_n} b^{(g)}_\mu v_{\mu_1} v_{\mu_2} \ldots v_{\mu_{\ell(\mu)}} \tag{42}$$

$$= \sum_{r=0}^{\infty} (2^r r!)^{g-1} \left(\tfrac{v(t)}{1+v(t)} \right)^r \tag{43}$$

by specializing Theorem 4.1 to commutative variables. This is a formal power series in the ring $\mathbb{Z}[v_1, v_2, \ldots][[t]]$.

The generating functions for counting *unsigned arrangements* by unordered type are

$$a_n^{(g)}(\vec{u}) = a_n^{(g)}(u_1, u_2, \ldots) = \omega_a(\mathcal{A}_n^{(g)}) = \sum_{\lambda \in \mathcal{P}_n} a_\lambda^{(g)} u_{\lambda_1} u_{\lambda_2} \ldots u_{\lambda_{\ell(\lambda)}} \tag{44}$$

$$a^{(g)}(\vec{u}; t) = a^{(g)}(u_1, u_2, \ldots; t) = \omega_a(\mathcal{A}^{(g)}; t) = \sum_{n=0}^{\infty} t^n \sum_{\lambda \in \mathcal{P}_n} a_\lambda^{(g)} u_{\lambda_1} u_{\lambda_2} \ldots u_{\lambda_{\ell(\lambda)}} \tag{45}$$

$$= \sum_{r=0}^{\infty} (2^r r!)^{g-1} \left(\frac{u(t)}{1+u(t)} + \widetilde{G} u_1 t \right)^r \tag{46}$$

by specializing Theorem 5.7 to commutative variables. This is a formal power series in the ring $\mathbb{Z}[u_1, u_2, \ldots][[t]]$.

The homomorphism ϕ of Section 5 induces homomorphisms $\phi : \mathbb{Q}[u_1, u_2, \ldots] \to \mathbb{Q}[v_1, v_2, \ldots]$ and $\phi : \mathbb{Q}[u_1, u_2, \ldots][[t]] \to \mathbb{Q}[v_1, v_2, \ldots][[t]]$ in the commutative case. We will see that these are isomorphisms. We summarize the results on formulas for ϕ:

Theorem 6.2. *(i) The unordered weight of all signages of unsigned* $\mathrm{id}_n^{(g)}$ *($n > 0$) is*

$$\phi(u_n) = \sum_{k=1}^{n} G^{k-1} {v_1}^{k-1} \sum_{\mu \in \mathcal{P}_{n+1,k}} M(\mu) v_{\mu_1 - 1, \mu_2 - 1, \ldots, \mu_k - 1} \tag{47}$$

$$= v_n + G v_1 \sum_{r=0}^{n-1} v_r \cdot \phi(u_{n-1-r}) \tag{48}$$

(ii) Let $\vec{\sigma}$ be an unsigned (n, g)-arrangement with unordered type λ. The unordered weight of all signages of $\vec{\sigma}$ is $\phi(u_{\lambda_1})\phi(u_{\lambda_2})\cdots$.

(iii) $\phi\big(a^{(g)}(\vec{u}; t)\big) = b^{(g)}(\vec{v}; t)$.

Proof. These follow by specializing Theorems 5.2, 5.4 and Corollary 5.3 to commutative variables. In (20), compositions with the same nonzero parts but in a different order result in identical terms once we allow the variables to commute; collecting like terms gives coefficient $M(\mu)$ in (47), for the number of such compositions. □

The analog of (22) is

$$\phi(u_\lambda) = \phi(u_{\lambda_1})\phi(u_{\lambda_2}) \cdots = \sum_{\mu \in \mathcal{P}_n} h_{\lambda\mu}(G) v_\mu \tag{49}$$

where we plug in (47), expand out the products and collect terms, and obtain transition matrix $h(G)$ from the coefficients. For $n \geq 0$, the matrix $h(G)$ is a $p(n) \times p(n)$ matrix (where $p(n)$ is the number of integer partitions of n), indexed by partitions $\lambda, \mu \in \mathcal{P}_n$. We list row λ and column μ indices in reverse lexicographic order on $\mathcal{P}_n$ (or any other extension of refinement order). If $\vec{\sigma}$ is an unsigned (n, g)-arrangement of unordered type λ, then $h_{\lambda\mu}(G)$ gives the number of signages of $\vec{\sigma}$ with unordered type μ.

Theorem 6.3. *All parts (i)–(vi) of Theorem 5.5 go through to the commutative case via the commutative specialization, with the following additional modifications:*

(ii) In refinement order on partitions, $h(G)$ is lower triangular with 1's on the diagonal.

(iii) $h(G)^{-1} = h(\widetilde{G})$. Thus $\phi^{-1}(v_\lambda) = \phi^{-1}(v_{\lambda_1})\phi^{-1}(v_{\lambda_2})\cdots = \sum_{\mu\in\mathcal{P}_n} h_{\lambda\mu}(\widetilde{G})u_\mu$.

7 Example: Unsigned arrangements counted by type

We will use the results of the preceding sections to explicitly compute $A^{(g)}_\alpha$, the number of unsigned (n,g)-arrangements with ordered type α, and $a^{(g)}_\lambda$, the number of unsigned (n,g)-arrangements with unordered type λ. Fix $n>0$. To compute $A^{(g)}_\alpha$ for all $\alpha\in\mathcal{C}_n$,

1. Compute the ordered weight of all signed (n,g)-arrangements,

$$B^{(g)}_n(\vec{V}) = \omega_B(\mathcal{B}^{(g)}_n) = \sum_{\beta\in\mathcal{C}_n} B^{(g)}_\beta V_\beta = \sum_{k=0}^{n} b^{(g)}_{k,k} \sum_{\beta\in\mathcal{C}_{n,k}} V_\beta$$

   where $b^{(g)}_{k,k}$ is given by (1), the double sum has 2^{n-1} terms, and $V_\beta = V_{\beta_1}V_{\beta_2}\cdots$.

2. Compute $A^{(g)}_n(\vec{U}) = \omega_A(\mathcal{A}^{(g)}_n) = \phi^{-1}(B^{(g)}_n(\vec{V}))$, the ordered weight of all unsigned (n,g)-arrangements. Use (24) to compute $\phi^{-1}(V_1),\ldots,\phi^{-1}(V_n)$, and use that ϕ^{-1} is multiplicative and additive.

3. Collect terms by monomials in the U's: $A^{(g)}_n(\vec{U}) = \sum_{\alpha\in\mathcal{C}_n} A^{(g)}_\alpha U_\alpha$. The coefficient of $U_\alpha = U_{\alpha_1}U_{\alpha_2}\cdots$ is $A^{(g)}_\alpha$.

We may compute $a^{(g)}_\lambda$ for all $\lambda\in\mathcal{P}_n$ via the corresponding commutative formulas. Or, use $a^{(g)}_\lambda = \sum_\alpha A^{(g)}_\alpha$, where α runs over distinct compositions obtained by permuting the parts of λ.

Table 2 shows all unsigned $(4,2)$-arrangements. An $(n,2)$-arrangement is $(\sigma^{(1)},\sigma^{(2)})$ where $\sigma^{(1)}$ is the identity, so the table only shows the values of $\sigma^{(2)}$. We will apply the above algorithm to compute the counts in this table for $(4,2)$-arrangements, and for $(4,g)$-arrangements for general g.

Let $g=2$, so $G=1$ and $\widetilde{G}=-\frac{1}{2}$. By (1), the number of incompressible signed $(k,2)$-arrangements for $k=1,2,3,4$ is $b^{(2)}_{1,1}=2$, $b^{(2)}_{2,2}=6$, $b^{(2)}_{3,3}=34$, $b^{(2)}_{4,4}=262$. By (24),

$$\begin{aligned}
\phi^{-1}(V_1) &= \tfrac{1}{2}U_1\\
\phi^{-1}(V_2) &= U_2 - \tfrac{3}{4}U_1U_1\\
\phi^{-1}(V_3) &= U_3 - \tfrac{1}{2}U_2U_1 - \tfrac{1}{2}U_1U_2 + \tfrac{1}{8}U_1U_1U_1\\
\phi^{-1}(V_4) &= U_4 - \tfrac{1}{2}U_3U_1 - \tfrac{1}{2}U_1U_3 - \tfrac{1}{4}U_2U_1U_1 + \tfrac{1}{4}U_1U_2U_1 - \tfrac{1}{4}U_1U_1U_2 + \tfrac{5}{16}U_1U_1U_1U_1
\end{aligned}$$

# strips k	$a^{(2)}_{4,k}$	**Unordered type λ**	$a^{(2)}_{\lambda}$	**Ordered type α**	$A^{(2)}_{\alpha}$	**Permutations $\pi^{(2)}$**
1	2	(4)	2	(4)	2	$\langle 1,2,3,4\rangle; \langle 4,3,2,1\rangle$
2	10	$(3,1)$	4	$(3,1)$	2	$\langle 3,2,1 \mid 4\rangle; \langle 4 \mid 1,2,3\rangle$
				$(1,3)$	2	$\langle 1 \mid 4,3,2\rangle; \langle 2,3,4 \mid 1\rangle$
		$(2,2)$	6	$(2,2)$	6	$\langle 1,2 \mid 4,3\rangle; \langle 2,1 \mid 3,4\rangle; \langle 2,1 \mid 4,3\rangle;$ $\langle 3,4 \mid 1,2\rangle; \langle 3,4 \mid 2,1\rangle; \langle 4,3 \mid 1,2\rangle$
3	10	$(2,1,1)$	10	$(2,1,1)$	2	$\langle 3 \mid 1,2 \mid 4\rangle; \langle 4 \mid 2,1 \mid 3\rangle$
				$(1,2,1)$	6	$\langle 1 \mid 3,2 \mid 4\rangle; \langle 1 \mid 4 \mid 2,3\rangle; \langle 2,3 \mid 1 \mid 4\rangle;$ $\langle 3,2 \mid 4 \mid 1\rangle; \langle 4 \mid 1 \mid 3,2\rangle; \langle 4 \mid 2,3 \mid 1\rangle$
				$(1,1,2)$	2	$\langle 1 \mid 3,4 \mid 2\rangle; \langle 2 \mid 4,3 \mid 1\rangle$
4	2	$(1,1,1,1)$	2	$(1,1,1,1)$	2	$\langle 2 \mid 4 \mid 1 \mid 3\rangle; \langle 3 \mid 1 \mid 4 \mid 2\rangle$

Table 2: Unsigned $(4,2)$-arrangements: $\vec{\pi} = (\pi^{(1)}, \pi^{(2)})$ on $n = 4$ elements with $g = 2$ permutations. $\pi^{(1)} = \langle 1,2,3,4\rangle$ = identity and $\pi^{(2)} \in \mathcal{S}_4$ as listed. $\pi^{(2)}$ is given in one-line permutation notation, but annotated with vertical bars between strips. There are $A^{(2)}_\alpha$ arrangements of ordered type α; $a^{(2)}_\lambda$ of unordered type λ; and $a^{(2)}_{4,k}$ with k strips.

The signed $(4,2)$-arrangements have weight

$$\begin{aligned}
B^{(2)}_4(\vec{V}) &= B^{(2)}_{(4)} V_4 + B^{(2)}_{(31)} V_3 V_1 + B^{(2)}_{(13)} V_1 V_3 + B^{(2)}_{(22)} V_2 V_2 \\
&\qquad + B^{(2)}_{(211)} V_2 V_1 V_1 + B^{(2)}_{(121)} V_1 V_2 V_1 + B^{(2)}_{(112)} V_1 V_1 V_2 + B^{(2)}_{(1111)} V_1 V_1 V_1 V_1 \\
&= b^{(2)}_{1,1} V_4 + b^{(2)}_{2,2} (V_3 V_1 + V_1 V_3 + V_2 V_2) \\
&\qquad + b^{(2)}_{3,3} (V_2 V_1 V_1 + V_1 V_2 V_1 + V_1 V_1 V_2) + b^{(2)}_{4,4} V_1 V_1 V_1 V_1 \\
&= 2V_4 + 6(V_3 V_1 + V_1 V_3 + V_2 V_2) \\
&\qquad + 34(V_2 V_1 V_1 + V_1 V_2 V_1 + V_1 V_1 V_2) + 262 V_1 V_1 V_1 V_1 \,. \qquad (50)
\end{aligned}$$

The unsigned $(4,2)$-arrangements have weight $A^{(2)}_4(\vec{U}) = \phi^{-1}(B^{(2)}_4(\vec{V}))$:

$$\begin{aligned}
A^{(2)}_4(\vec{U}) &= \phi^{-1}(B^{(2)}_4(\vec{V})) \\
&= 2\phi^{-1}(V_4) + 6\left(\phi^{-1}(V_3)\phi^{-1}(V_1) + \phi^{-1}(V_1)\phi^{-1}(V_3) + \phi^{-1}(V_2)\phi^{-1}(V_2)\right) + \cdots \\
&= 2U_4 + 2U_3 U_1 + 2U_1 U_3 + 6U_2 U_2 \\
&\qquad + 2U_2 U_1 U_1 + 6U_1 U_2 U_1 + 2U_1 U_1 U_2 + 2U_1 U_1 U_1 U_1 \,. \qquad (51)
\end{aligned}$$

For each $\alpha \in \mathcal{C}_4$, the coefficient of $U_{\alpha_1} U_{\alpha_2} \cdots$ is $A^{(2)}_\alpha$. Compare with Table 2.

For unordered types, replace $U_i \to u_i$, $V_i \to v_i$ so that they commute. This gives

$$\begin{aligned}
b^{(2)}_4(\vec{v}) &= 2v_4 + 12 v_3 v_1 + 6{v_2}^2 + 102 v_2 {v_1}^2 + 262 {v_1}^4 \\
a^{(2)}_4(\vec{u}) &= 2u_4 + 4u_3 u_1 + 6u_2 u_2 + 10 u_2 {u_1}^2 + 2{u_1}^4
\end{aligned}$$

This gives the $a_\lambda^{(2)}$ column in Table 2. Finally, the generating function for the number of strips is obtained by specializing $v_i \to z$ or $u_i \to z$, and is distinguished notationally by having a scalar argument z instead of a vector argument. $a_4^{(2)}(z)$ gives the $a_{4,k}^{(2)}$ column.

$$\begin{aligned} b_4^{(2)}(z) &= 2z + 18z^2 + 102z^3 + 262z^4 \\ a_4^{(2)}(z) &= 2z + 10z^2 + 10z^3 + 2z^4 \end{aligned}$$

In the above computations, $g = 2$ gave $G = 1$ and $\widetilde{G} = -\frac{1}{2}$. To compute $A_\alpha^{(g)}$ for all g, we leave g as a variable, so that $G = 2^{g-1} - 1$ and $\widetilde{G} = 2^{-(g-1)} - 1$. By (1), the number of incompressible signed (n, g)-arrangements for $n = 1, 2, 3, 4$ is

$$\begin{aligned} b_{1,1}^{(g)} &= 2^{g-1} & \qquad b_{3,3}^{(g)} &= 48^{g-1} - 2(8^{g-1}) + 2^{g-1} \\ b_{2,2}^{(g)} &= 8^{g-1} - 2^{g-1} & \qquad b_{4,4}^{(g)} &= 384^{g-1} - 3(48^{g-1}) + 3(8^{g-1}) - 2^{g-1} \end{aligned}$$

which replace $2, 6, 34, 262$ in (50). The map from signed weights to unsigned weights is

$$\begin{aligned} \phi^{-1}(V_1) &= (\widetilde{G} + 1)U_1 \\ \phi^{-1}(V_2) &= U_2 + \widetilde{G}(\widetilde{G} + 2)U_1U_1 \\ \phi^{-1}(V_3) &= U_3 + \widetilde{G}U_2U_1 + \widetilde{G}U_1U_2 + \widetilde{G}(\widetilde{G}^2 + 3\widetilde{G} + 1)U_1U_1U_1 \\ \phi^{-1}(V_4) &= U_4 + \widetilde{G}U_3U_1 + \widetilde{G}U_1U_3 + \widetilde{G}(\widetilde{G} + 1)U_2U_1U_1 \\ &\qquad + \widetilde{G}^2U_1U_2U_1 + \widetilde{G}(\widetilde{G} + 1)U_1U_1U_2 + \widetilde{G}^2(\widetilde{G} + 3)(\widetilde{G} + 1)U_1U_1U_1U_1 \end{aligned}$$

We ultimately arrive at these counts of unsigned $(4, g)$-arrangements by ordered type:

$$\begin{aligned} A_{(4)}^{(g)} &= 2^{g-1} & \qquad A_{(2,1,1)}^{(g)} = A_{(1,1,2)}^{(g)} &= 12^{g-1} - 4^{g-1} - 8^{g-1} + 2^{g-1} \\ A_{(3,1)}^{(g)} = A_{(1,3)}^{(g)} &= 4^{g-1} - 2^{g-1} & \qquad A_{(1,2,1)}^{(g)} &= 12^{g-1} - 2(4^{g-1}) + 2^{g-1} \\ A_{(2,2)}^{(g)} &= 8^{g-1} - 2^{g-1} & \qquad A_{(1,1,1,1)}^{(g)} &= 24^{g-1} - 3(12^{g-1}) + 8^{g-1} + 2(4^{g-1}) - 2^{g-1} \end{aligned}$$

and these counts by unordered type:

$$\begin{aligned} a_{(4)}^{(g)} &= 2^{g-1} & \qquad a_{(2,1,1)}^{(g)} &= 3(12^{g-1}) - 2(8^{g-1}) - 4(4^{g-1}) + 3(2^{g-1}) \\ a_{(3,1)}^{(g)} &= 2(4^{g-1}) - 2(2^{g-1}) & \qquad a_{(1,1,1,1)}^{(g)} &= 24^{g-1} - 3(12^{g-1}) + 8^{g-1} + 2(4^{g-1}) - 2^{g-1} \\ a_{(2,2)}^{(g)} &= 8^{g-1} - 2^{g-1} \end{aligned}$$

8 Applications to rearrangement distance and conserved segments

In genome studies using unsigned marker data, it is common to use the canonical signage; although this may not properly resolve the signs of the singletons, it does establish strip boundaries. However, there could be undetected flips of individual markers within the strips. In Section 8.1, we will see that for two genomes under the uniform distribution, the canonical signage leads to errors in strip boundaries in $\approx 77\%$ of all cases. In Section 8.2, we will study a manifestation of this error in a synteny block detection algorithm by Sankoff and Trinh [29, 30]. In Section 8.3, we will study the number of arrangements when a minimum or maximum strip length is imposed (for example, to filter out singletons).

In Section 8.4, we will describe issues and potential future work concerning the difference in the distribution of incompressible arrangements (typically representing segment orders) vs. arbitrary arrangements (typically representing gene orders).

8.1 Incorrect identification of conserved segments due to misclassified signs

We consider genome rearrangement studies that determine conserved segments as strips in unsigned marker data. The following theorem shows that if all arrangements are equally likely, the canonical signage is likely to make errors in determining strip boundaries for $\approx 77\%$ of all cases with two genomes when n is large, but is unlikely to make errors in the boundaries for three or more genomes when n is large. We are only addressing the strip boundaries; the signs of singleton elements remain ambiguous, but changing signs of singletons does not affect strip boundaries.

Theorem 8.1. *Let $\vec{\sigma}$ range over $\mathcal{B}_{n,n}^{(g)}$. As $n \to \infty$, the probability that $|\vec{\sigma}|$ has fewer than n unsigned strips approaches $1 - \exp(-3/2) \approx 0.7769$ if $g = 2$, and approaches 0 if $g > 2$.*

Proof. Let $\vec{\sigma} \in \mathcal{B}_n^{(g)}$ and consider the unsigned arrangement $|\vec{\sigma}|$. The number of strips in $|\vec{\sigma}|$ is less or equal to the number of strips in σ, so if $|\vec{\sigma}|$ has n unsigned strips then $\vec{\sigma}$ must have n signed strips. (The converse need not hold.)

Thus, the number of arrangements $\vec{\sigma} \in \mathcal{B}_{n,n}^{(g)}$ with $|\vec{\sigma}| \in \mathcal{A}_{n,n}^{(g)}$ is $2^{n(g-1)} a_{n,n}^{(g)}$ (by assigning all possible signs to the n elements in all but the first genome).

The number of arrangements $\vec{\sigma} \in \mathcal{B}_{n,n}^{(g)}$ with $|\vec{\sigma}| \notin \mathcal{A}_{n,n}^{(g)}$ is $b_{n,n}^{(g)} - 2^{n(g-1)} a_{n,n}^{(g)}$.

The fraction of arrangements $\vec{\sigma} \in \mathcal{B}_{n,n}^{(g)}$ for which $|\vec{\sigma}| \notin \mathcal{A}_{n,n}^{(g)}$ is

$$P\left(|\vec{\sigma}| \notin \mathcal{A}_{n,n}^{(g)} \;\middle|\; \vec{\sigma} \in \mathcal{B}_{n,n}^{(g)}\right) = \frac{b_{n,n}^{(g)} - 2^{n(g-1)} a_{n,n}^{(g)}}{b_{n,n}^{(g)}} = 1 - \frac{a_{n,n}^{(g)}/n!^{g-1}}{b_{n,n}^{(g)}/(2^{n(g-1)} n!^{g-1})}\,. \tag{52}$$

For 2 genomes, the $g = 2$, $q = 0$ cases of (40) and (5) show that this approaches $1 - \exp(-2)/\exp(-\frac{1}{2}) = 1 - \exp(-3/2)$ as $n \to \infty$. See Fig. 2(c).

For $g > 2$ genomes, the $q = 0$ cases of (39) and (4) are

$$\lim_{n\to\infty} \frac{a_{n,n}^{(g)}}{n!^{g-1}} = 1 \qquad \text{and} \qquad \lim_{n\to\infty} \frac{b_{n,n}^{(g)}}{2^{n(g-1)} n!^{g-1}} = 1$$

so (52) approaches 0 as $n \to \infty$. □

Notes:

1. Genome rearrangement studies typically assume that genomes evolved through a series of rearrangement events, such as a certain number of reversals. Assuming that the arrangements are due to a particular number of reversals leads to a non-uniform probability distribution over all arrangements; Theorem 8.1 assumes a uniform probability distribution over all arrangements, so it would not apply in that situation.

2. Sankoff and Trinh [29, 30] introduced an algorithm for constructing syntenic blocks for two genomes, that was intended to be simpler to analyze than the algorithm by Pevzner and Tesler [25]. Peng, Pevzner and Tesler [23] determined a number of flaws in the Sankoff-Trinh construction, one of which is quantified by this theorem. This will be described in the next section.

8.2 Sankoff and Trinh: Synteny block construction

In a debate between Pevzner and Tesler [24] and Sankoff and Trinh [29, 30] concerning the random breakage model of evolution, Sankoff and Trinh introduced a synteny block construction algorithm. Peng et al. [23] found a number of flaws in this algorithm, and stated that a certain error would occur in this algorithm in approximately 77% of all signed permutations. We will now prove that; for full details on the context, see [23]. The portion of Sankoff and Trinh's algorithm relevant to this discussion is as follows. Let π be a signed permutation on $1, \dots, n$. There are also integer parameters $w \geq 0$ and $\Delta \geq 1$. In our terminology, this concerns the signed $(n, 2)$-arrangement (id_n, π).

ST-Synteny(π, w, Δ)

Step 1: Define each element of π as a block and iteratively amalgamate the resulting blocks as follows: two adjacent blocks in π are amalgamated if for some $i, j > 0$, element i or $-i$ is in one block and j or $-j$ is in the other block and $|i - j| \leq w$. Signs of the elements are recorded but ignored during amalgamation.

Step 2: Delete any "short" block containing less than Δ elements ($\Delta = 3$ in [29]).

We recount an example from [23] of one of the types of errors that occurs with this algorithm. Given one reversal

$$\begin{array}{rccc} \sigma^{(1)}: & 1 \dots 100 & \underline{101 \dots 200} & 201 \dots 300 \\ \pi = \sigma^{(2)}: & 1 \dots 100 & -200 \dots -101 & 201 \dots 300 \end{array} \tag{53}$$

there should be 3 blocks and 2 breakpoints. However, there is no setting of the parameters of ST-Synteny that does this. If $w = 0$, the 300 separate elements become 300 separate blocks. These would be deleted for being too small with any setting of the minimum block size $\Delta > 1$, so we set $\Delta = 1$. If $w \geq 1$, ST-Synteny forms one large block, as follows. First it does a number of amalgamation steps resulting in the expected 3 blocks:

$$\langle 1 \dots 100 \rangle \quad \langle -200 \dots -101 \rangle \quad \langle 201 \dots 300 \rangle \tag{54}$$

Essentially to this point, we have performed unsigned compression on $|\pi|$ (the bijection in Theorem 2.4; unsigned because Step 1 ignores the signs of elements).

But the iterative amalgamation does not stop there. Since 100 and -101 are now in adjacent blocks and $|101-100| = 1 \le w$, it merges the first two blocks. Since -200 and 201 are in adjacent blocks, those blocks are also merged. The result is a single block

$$\langle 1 \ldots 100 \quad -200 \ldots -101 \quad 201 \ldots 300 \rangle \tag{55}$$

In particular, there is no setting of the ST-Synteny parameters that will produce strips as defined in this paper, although that should be what $w = \Delta = 1$ produces. Instead, the setting $\Delta = w = 1$ is equivalent to repeatedly applying absolute value and unsigned compression in any order, until the result stabilizes. On the other hand, iterating signed compression would stabilize in one step. We now explain this more formally.

Definition 8.2 (Signed compression of a signed permutation). For a signed permutation $\pi \in \mathcal{B}_n$, consider the signed $(n,2)$-arrangement (id_n, π). Compute the signed compression $\Psi_b(\mathrm{id}_n, \pi) = (\vec{\tau}, \beta)$ (Theorem 2.2), where $\vec{\tau} = (\tau^{(1)}, \tau^{(2)})$ is a signed arrangement and β is a composition. Define $f_b(\pi) = \tau^{(2)}$.

Definition 8.3 (Unsigned compression of a permutation). For a permutation $\pi \in \mathcal{S}_n$ or signed permutation $\pi \in \mathcal{B}_n$, form the unsigned $(n,2)$-arrangement $(\mathrm{id}_n, |\pi|)$. Compute the unsigned compression $\Psi_a(\mathrm{id}_n, |\pi|) = (\vec{\tau}, \alpha)$ (Theorem 2.4), where $|\pi|$ means to take the absolute value of each element in π. Again, $\vec{\tau} = (\tau^{(1)}, \tau^{(2)})$ is a signed arrangement and α is a composition. Define $f_a(\pi) = \tau^{(2)}$.

For $w = \Delta = 1$, the ST-Synteny algorithm reduces to repeated applications of f_a to π until it stabilizes. If $f_a(f_b(\pi))$ has the same size as $f_b(\pi)$, then this algorithm stabilizes in one step and the result is a correct decomposition into strips. Otherwise, the algorithm overamalgamates.

Continuing with the example above, the signed compression of π in (53) is $f_b(\pi) = \langle 1, -2, 3 \rangle$, where "1" represents the compression of $1, \ldots, 100$; "-2" represents the compression of $-200, \ldots, -101$; and "3" represents the compression of $201, \ldots, 300$. Signed compression is stable in one step, since $f_b(f_b(\pi)) = f_b(\pi)$. Unsigned compression also gives $f_a(\pi) = \langle 1, -2, 3 \rangle$ (cf. (54)), but the orientation is ignored so it undergoes further compression: $f_a(f_a(\pi)) = f_a(\langle 1, -2, 3 \rangle) = f_a(\langle 1, 2, 3 \rangle) = \langle 1 \rangle$ (cf. (55)). Then it terminates since $f_a(\langle 1 \rangle) = \langle 1 \rangle$ does not reduce the length. In $\langle 1 \rangle$, "1" represents the whole original permutation, without indication of the details of the rearrangements therein.

Next, we quantify how often over-amalgamation occurs. Consider signed permutations on $n \ge 1$ elements with $k \ge 1$ strips, $\pi \in \mathcal{B}_{n,k}$. There are $b^{(2)}_{n,k} = \binom{n-1}{k-1} b^{(2)}_{k,k}$ such permutations, of which $\binom{n-1}{k-1} 2^k a^{(2)}_{k,k}$ would also compress to exactly k strips using f_a instead of f_b. Thus the total number of signed permutations of size n that are correctly broken into strips by this algorithm is

$$\text{(a)} \ \sum_{k=1}^{n} \binom{n-1}{k-1} 2^k a^{(2)}_{k,k} \qquad \text{out of} \qquad \text{(b)} \ \sum_{k=1}^{n} \binom{n-1}{k-1} b^{(2)}_{k,k} = 2^n \, n! \tag{56}$$

total signed arrangements. The ratio of the kth term of (56a) to the kth term of (56b) approaches $\exp(-3/2)$ for large k as $n \to \infty$, by the proof of Theorem 8.1. The terms grow fast enough by (5) and (40) that the ratio of sum (56a) to (56b) approaches $\exp(-3/2)$.

The fraction of signed arrangements of size n that are not correctly broken into strips by this algorithm therefore approaches $1 - \exp(-3/2)$.

8.3 Minimum and maximum strip length

In [13], Pevzner and Hannenhalli give an algorithm for assigning signs in unsigned arrangements with two genomes, based on an optimization criterion (minimizing the signed reversal distance between the genomes). Under their optimization criterion, singletons will only have the canonical signage half the time (canonical for singletons is positive); strips of length 2 usually have the canonical signage, but not always; and strips of length 3 or greater are always assigned the canonical signage. Their optimization criterion is useful computationally, although evolution did not necessarily optimize their criterion. This motivates the criteria that in an unsigned genome, signs should be assigned canonically in all unsigned strips of length $\geq L$, while unsigned strips of length $< L$ may have ambiguous signs and should be deleted.

Theorem 8.4. *Let $L > 1$. The generating function for the number $a^{(g)}_{\geq L}(n)$ of unsigned (n, g)-arrangements in which all strips have length $\geq L$ is*

$$\sum_{n=0}^{\infty} a^{(g)}_{\geq L}(n)\, t^n = \sum_{r=0}^{\infty} r!^{g-1} \left(\frac{2^{g-1} t^L}{1 - t + t^L} \right)^r = b^{(g)}(t, t^{L-1}) \,. \tag{57}$$

Proof. Specialize (46) by setting $u_i = 0$ for $i < L$, $u_i = 1$ for $i \geq L$.

Then $\widetilde{G}u_1 t = 0$ ($u_1 = 0$ since $L > 1$) and $u(t) = \sum_{i=L}^{\infty} t^i = t^L/(1-t)$.

Plug these into (46) and simplify to get the left side of (57).

Observe that plugging $z = t^{L-1}$ into (13) gives that same summation, hence we have the right side of (57). A combinatorial interpretation is that unsigned (n, g)-arrangements with k parts, all of length $\geq L$, are in a bijection with signed $(n - k(L-1), g)$-arrangements with k parts, all of length ≥ 1. Starting with such an unsigned arrangement, apply the canonical signage, remove $L - 1$ entries from each strip, and renumber the remaining elements to obtain such a signed arrangement. This is reversible. $\square$

Note that Abramson and Moser, 1967 [1, p. 1249, Eq. (ii)] have a multiple summation formula ($L - 2$ nested sums) for the $g = 2$ case of this.

In studying the reversal distance between two genomes and phylogenetic trees based on multiple genomes, in the signed case one may compress all strips to singletons without affecting the distances. In the unsigned case, by the Pevzner and Hannenhalli [13] result, one may retain strips of length ≤ 3, and compress each strip of length ≥ 4 to a strip of length 3, without affecting the distances. We may specialize (46) and (43) to obtain generating functions that count arrangements by maximum strip length:

Theorem 8.5. *Let $L > 1$. Let $a^{(g)}_{<L}(n)$ and $b^{(g)}_{<L}(n)$ respectively be the numbers of unsigned and signed (n, g)-arrangements in which all strips have length $< L$. Then*

$$\sum_{n=0}^{\infty} a^{(g)}_{<L}(n) t^n = \sum_{r=0}^{\infty} (2^r\, r!)^{g-1} \left(\tfrac{t-t^L}{1-t^L} + \widetilde{G}t \right)^r , \qquad \sum_{n=0}^{\infty} b^{(g)}_{<L}(n) t^n = \sum_{r=0}^{\infty} (2^r\, r!)^{g-1} \left(\tfrac{t-t^L}{1-t^L} \right)^r .$$

8.4 Rearrangement distance for segments vs. genes

Some papers work with conserved segments while others work with conserved genes. The distribution of incompressible permutations, signed or unsigned, is different than the distribution of all permutations.

The *reversal distance* between two unichromosomal genomes is the minimum number of reversals required to convert one into the other. For two signed genomes, a polynomial-time algorithm for it was first found by Pevzner and Hannenhalli [14], and a linear time algorithm was developed by Bader, Moret, and Yan, 2001 [4].

For n genes, Kececioglu and Sankoff, 1994 [17] and Bourque and Pevzner, 2002 [8] used simulations to empirically compute the expected reversal distance d to the identity after performing r random signed reversals on the identity. [17] determined that $E(d) \approx r$ for $r \leq .5n$ when $n = 1000$, and [8] determined $E(d) \approx r$ for $r \leq .4n$ when $n = 100$. Both of these studies were based on simulations. Berestycki and Durrett [7] used mathematical analysis (rather than simulations) to show that $E(d) \approx r$ for $r \leq .5n$ as $n \to \infty$, but they do not quantify how large n has to be for this to be practical. Xu et al., 2006 [35] compute the expected rearrangement distance between two multichromosomal genomes with a specified number of genes and chromosomes, using additional rearrangement operations that may apply to multichromosomal genomes.

These studies assume r random rearrangement operations are performed on n genes, and analyze the minimal number of rearrangements $d \leq r$ that could have produced that arrangement. The number of breakpoints depends on the particular details of the rearrangements. However, genome rearrangement studies that start by computing strips (or more generally conserved segments or syntenic blocks; see [19, 24]) will result in $k \leq n$ strips, with $k - 1$ internal breakpoints. Results that analyze the relation between d and r based on n genes do not directly apply to the relation between d and r based on k segments, particularly if $k \ll n$. Future work should address this; [23, 29, 30] are a start.

9 Circular chromosomes

The arrangements described in previous sections of this paper are *linear arrangements*, which model unichromosomal linear genomes. In this section, we will derive analogous results for *circular arrangements*, which model unichromosomal circular genomes. Section 9.1 gives an example of a circular arrangement, defines the notation, and gives basic enumeration results for counting the number of signed circular arrangements with specified type or number of strips. In Section 9.2, we derive the generating functions for the number of signed circular arrangements by ordered type, and in Section 9.3 we do the same for unsigned circular arrangements. Section 9.4 gives the generating functions for unordered types. Finally, in Section 9.5, we specialize these results to get generating functions for the number of strips and number of incompressible arrangements.

9.1 Circular arrangements, notation and example

Unsigned circular permutations of $\{1,\ldots,n\}$ are unsigned permutations $\langle j_1,\ldots,j_n\rangle$ of $\{1,\ldots,n\}$ modulo cyclic shifts:

$$\langle j_1,\ldots,j_n\rangle \equiv \langle j_2,\ldots,j_n,j_1\rangle \equiv \cdots \equiv \langle j_n,j_1,\ldots,j_{n-1}\rangle \tag{58}$$

and reversing the whole sequence:

$$\langle j_1,\ldots,j_n\rangle \equiv \langle j_n,j_{n-1},\ldots,j_1\rangle \tag{59}$$

For $n \geq 3$ there are $2n$ equivalent linear representations of each unsigned circular permutation, while for $n < 3$ some reversals and cyclic shifts are the same. Let $\mathring{\mathcal{S}}_n$ be the set of unsigned circular permutations of $\{1,\ldots,n\}$. Then $|\mathring{\mathcal{S}}_n| = 1$ for $n = 0,1,2,3$ and $|\mathring{\mathcal{S}}_n| = (n-1)!/2$ for $n \geq 3$. Technically, the action just described is dihedral rather than circular, but the term *circular genome* is well-established.

Signed circular permutations of $\{1,\ldots,n\}$ are signed permutations $\langle j_1,\ldots,j_n\rangle$ of $\{1,\ldots,n\}$ modulo cyclic shifts (58) and negatives in place of (59):

$$\langle j_1,j_2,\ldots,j_n\rangle \equiv \langle -j_n,\ldots,-j_2,-j_1\rangle \tag{60}$$

For $n \geq 1$ there are $2n$ equivalent linear representations of each signed circular permutation. Let $\mathring{\mathcal{B}}_n$ be the set of signed circular permutations on $1,\ldots,n$. Then $|\mathring{\mathcal{B}}_n| = 1$ for $n = 0$ and $2^{n-1}(n-1)!$ for $n \geq 1$.

We will treat the case of $g \geq 2$ circular genomes at a time, corresponding to a tuple of g circular permutations in which the first is the identity. An *(un)signed circular* (n,g)-*arrangement* is a g-tuple of permutations $\vec{\sigma} = (\sigma^{(1)},\ldots,\sigma^{(g)})$ in $\mathring{\mathcal{S}}_n$ (unsigned) or $\mathring{\mathcal{B}}_n$ (signed) where $\sigma^{(1)} = \langle 1,\ldots,n\rangle$. Let $\mathring{\mathcal{A}}_n^{(g)}$ denote the set of unsigned circular (n,g)-arrangements and $\mathring{\mathcal{B}}_n^{(g)}$ denote the set of signed circular arrangements. Let $\mathring{\mathcal{A}}^{(g)}$ and $\mathring{\mathcal{B}}^{(g)}$ denote the sets of (un)signed circular arrangements of all sizes.

Adjacencies and breakpoints are defined as for linear arrangements, except that consecutive entries of the permutations now are modulo the relations described above (so the sequences do not have starts/ends but instead wrap around; thus, entries $(n,1)$ are consecutive in $\sigma^{(1)}$). *Linear strips* are sequences of consecutive entries of $\sigma^{(1)}$ bounded on both sides by breakpoints. *Circular strips* arise when there are no breakpoints. The term *strip* alone will refer to linear strips. Further technicalities will be described in Note 9.2.

Example 9.1. Consider this signed circular $(12,3)$-arrangement:

$$\begin{array}{ll}
\sigma^{(1)}: & \langle 1,2,3,4,5,6,7,8,9,10,11,12\rangle \\
\sigma^{(2)}: & \langle 3,4,5,11,12,1,-10,-9,-8,-7,-6,2\rangle \\
\sigma^{(3)}: & \langle 9,10,-4,-3,-2,-1,-12,11,-7,-6,-5,8\rangle
\end{array}$$

A different but equivalent representation of this same arrangement is

$$\begin{array}{ll} \sigma^{(1)}: & \langle 12, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\rangle \\ \sigma^{(2)}: & \langle 12, 1, -10, -9, -8, -7, -6, 2, 3, 4, 5, 11\rangle \\ \sigma^{(3)}: & \langle 12, 1, 2, 3, 4, -10, -9, -8, 5, 6, 7, -11\rangle \end{array}$$

The breakpoints in $\sigma^{(1)}$ are $(1, 2), (4, 5), (5, 6), (7, 8), (10, 11), (11, 12)$, and the adjacencies are $(12, 1), (2, 3), (3, 4), (6, 7), (8, 9), (9, 10)$. We code $(i-1, i)$ as i (for $i = 2, \ldots, n$) and $(n, 1)$ as 1, giving breakpoint set $S = \{2, 5, 6, 8, 11, 12\}$. The 6 breakpoints break this arrangement into 6 linear strips:

$$\begin{array}{lll} \sigma^{(1)}: & \boxed{12, 1}, \boxed{2, 3, 4}, \boxed{5}, \boxed{6, 7}, \boxed{8, 9, 10}, \boxed{11} & = J_1, J_2, J_3, J_4, J_5, J_6 \\ \sigma^{(2)}: & \boxed{12, 1}, \boxed{-10, -9, -8}, \boxed{-7, -6}, \boxed{2, 3, 4}, \boxed{5}, \boxed{11} & = J_1, J_2, -J_5, J_3, J_4, -J_6 \\ \sigma^{(3)}: & \boxed{12, 1}, \boxed{2, 3, 4}, \boxed{-10, -9, -8}, \boxed{5}, \boxed{6, 7}, \boxed{-11} & = J_1, J_2, -J_5, J_3, J_4, -J_6 \end{array}$$

The strips are $J_1 = \langle 12, 1\rangle$, $J_2 = \langle 2, 3, 4\rangle$, $J_3 = \langle 5\rangle$, $J_4 = \langle 6, 7\rangle$, $J_5 = \langle 8, 9, 10\rangle$, and $J_6 = \langle 11\rangle$, where J_1 is the strip in $\sigma^{(1)}$ with 1 and the strips are numbered consecutively in $\sigma^{(1)}$ from there. The *ordered 1-type* of the arrangement is the lengths of the consecutive strips in $\sigma^{(1)}$ starting with J_1: $\gamma = (2, 3, 1, 2, 3, 1)$. The *ordered type* is the lexicographically largest cyclic shift of the 1-type, $\beta = (3, 1, 2, 3, 1, 2)$, and the ordered weight is based on this, $V_3V_1V_2V_3V_1V_2$ (noncommuting V_i's). The *unordered type* is the partition formed by putting the strip lengths in weakly decreasing order, $\mu = (3, 3, 2, 2, 1, 1)$, and the unordered weight is ${v_3}^2{v_2}^2{v_1}^2$ (commuting v_i's).

Replacing $\pm J_i$ by $\pm i$ gives the *compression* of $\vec{\sigma}$ as a signed circular arrangement:

$$\begin{array}{ll} \tau^{(1)}: & \langle 1, 2, 3, 4, 5, 6\rangle \\ \tau^{(2)}: & \langle 1, -5, -4, 2, 3, 6\rangle \\ \tau^{(3)}: & \langle 1, 2, -5, 3, 4, -6\rangle \end{array}$$

Let $\mathring{\mathcal{C}}_n$ be the set of compositions of n that are lexicographically largest among all their cyclic shifts, and $\mathring{\mathcal{C}}_{n,k}$ be the subset of those which have k parts. Ordered 1-types range over $\mathcal{C}_n$; ordered types range over $\mathring{\mathcal{C}}_n$; and unordered types range over $\mathcal{P}_n$. The identity arrangement is an exception to this, and will be described in Note 9.2.

The *period* $\mathrm{per}(\gamma)$ of a composition $\gamma = (\gamma_1, \ldots, \gamma_k) \in \mathcal{C}_n$ is the minimum d $(1 \le d \le k)$ for which $(\gamma_1, \ldots, \gamma_k) = (\gamma_{d+1}, \ldots, \gamma_k, \gamma_1, \ldots, \gamma_d)$. The period divides $k = \ell(\gamma)$.

For $n > 1$, a signed circular (n, g)-arrangement (besides the identity) is *incompressible* if it equals its compression (equivalent to any of: it has no adjacencies; all its strips are singletons; its type is (1^n)). Note that the compression of any signed circular arrangement is incompressible. The circular identity arrangement has special conventions.

Note 9.2. Special conventions for the circular identity arrangement.

(i) Circular arrangements (both signed and unsigned) with $k > 0$ breakpoints break into k *linear strips*. It is impossible to have exactly 1 breakpoint. The circular identity arrangement $\mathrm{id}_n^{(g)}$ (all g permutations equal $\langle 1, \ldots, n\rangle$) has 0 breakpoints; we regard it as 0 *linear strips* and 1 *circular strip*. With this convention, circular arrangements with k breakpoints have k linear strips under all conditions.

(ii) As a signed circular arrangement, $\mathrm{id}_n^{(g)}$ (for $n \geq 0$) has type denoted by a new symbol, $C(n)$ (it is *not* (n), which would mean a linear strip of length n); ordered weight $\mathring{V}_n$; and unordered weight $\mathring{v}_n$. The variables with ring accents are new variables distinct from the non-ringed versions. As an unsigned circular arrangement, it has type $C(n)$, ordered weight $\mathring{U}_n$ and unordered weight $\mathring{u}_n$.

(iii) For circular signed and unsigned arrangements, the null arrangement $\mathrm{id}_0^{(g)}$ $(n = 0)$ is incompressible, while $\mathrm{id}_n^{(g)}$ for $n > 0$ is compressible with compression $\mathrm{id}_0^{(g)}$.

Let $n \geq 0$. Let $\mathring{\mathcal{B}}_{n,k}^{(g)}$ be the subset of $\mathring{\mathcal{B}}_n^{(g)}$ consisting of arrangements that break into $k \geq 0$ linear strips, and $\mathring{b}_{n,k}^{(g)} = |\mathring{\mathcal{B}}_{n,k}^{(g)}|$ be the number of such arrangements.

Let $\mathring{B}_\beta^{(g)}$ be the number of signed circular arrangements with ordered type $\beta \in \mathring{\mathcal{C}}_n$ and $\mathring{b}_\mu^{(g)}$ be the number of signed circular arrangements with unordered type $\mu \in \mathcal{P}_n$.

For $S \subset \{1, \ldots, n\}$, let $\mathring{\mathcal{B}}_{n,S}^{(g)}$ be the subset of $\mathring{\mathcal{B}}_n^{(g)}$ in which the breakpoint set is S. Let $\mathring{b}_{n,S}^{(g)} = |\mathring{\mathcal{B}}_{n,S}^{(g)}|$ be the number of such arrangements.

Theorem 9.3. *For $n \geq 0$ and $S \subset \{1, \ldots, n\}$ with $|S| = k > 0$, the procedure illustrated in Example 9.1 gives a bijection* ("compression")

$$\Psi_b \;:\; \mathring{\mathcal{B}}_{n,S}^{(g)} \to \mathring{\mathcal{B}}_{k,k}^{(g)}$$

between signed circular (n, g)-arrangements with k strips whose breakpoint set is S, and incompressible signed circular (k, g)-arrangements. For the circular identity arrangement, $\Psi_b(\mathrm{id}_n^{(g)}) = \mathrm{id}_0^{(g)}$, which gives the bijection for $n \geq 0$, $S = \emptyset$, $k = 0$.

Corollary 9.4. *(i) $\mathring{b}_{0,0}^{(g)} = 1$ (the null arrangement) and for $k > 0$,*

$$\mathring{b}_{k,k}^{(g)} = (-1)^k + \sum_{r=1}^{k} (-1)^{k-r} \binom{k}{r} (2^{r-1}(r-1)!)^{g-1} \,. \tag{61}$$

(ii) For all $S \subset \{1, \ldots, n\}$ with $|S| = k$, we have $\mathring{B}_{n,S}^{(g)} = \mathring{b}_{k,k}^{(g)}$.

(iii) $\mathring{b}_{n,k}^{(g)} = \mathring{b}_{k,k}^{(g)} \binom{n}{k}$.

(iv) For $\beta \in \mathring{\mathcal{C}}_{n,k}$, $\mathring{B}_\beta^{(g)} = \frac{n}{\ell(\beta)/\operatorname{per}(\beta)} \mathring{b}_{k,k}^{(g)}$. Note $\ell(\beta) = k$.

Due to conventions on the circular identity arrangement (Note 9.2), $\mathring{B}_{C(n)}^{(g)} = 1$ for $n \geq 0$ (extending the $k = 0$ case) and $\mathring{B}_{(n)}^{(g)} = 0$ for $n > 0$ (the $k = 1$ case).

(v) For $\mu \in \mathcal{P}_{n,k}$, $\mathring{b}_\mu^{(g)} = \mathring{M}(\mu) \mathring{b}_{k,k}^{(g)}$ where the number of subsets of $\{1, \ldots, n\}$ with unordered circular type μ is

$$\mathring{M}(\mu) = \frac{n}{\ell(\mu)} M(\mu) = \frac{n}{\ell(\mu)} \binom{\ell(\mu)}{m_1(\mu), m_2(\mu), \ldots} = \frac{n \cdot (\ell(\mu) - 1)!}{m_1(\mu)!\, m_2(\mu)! \cdots} \,. \tag{62}$$

Note $\ell(\mu) = k$. As in (iv), $\mathring{b}_{C(n)}^{(g)} = 1$ for $n \geq 0$ and $\mathring{b}_{(n)}^{(g)} = 0$ for $n > 0$.

Proof. The bijection in Theorem 9.3 gives (ii). Letting S range over all k-element subsets of $\{1,\ldots,n\}$ gives (iii). Letting S range over subsets of $\{1,\ldots,n\}$ of all sizes $k=0,\ldots,n$ gives the following, which will lead to (i):

$$\sum_{k=0}^{n} \mathring{b}_{k,k}^{(g)} \binom{n}{k} = |\mathring{\mathcal{B}}_n^{(g)}| = \begin{cases} (2^{n-1}(n-1)!)^{g-1} & \text{if } n \geq 1; \\ 1 & \text{if } n = 0. \end{cases} \tag{63}$$

This system of equations (63) for $n \geq 0$ may be solved for $\mathring{b}_{k,k}^{(g)}$, giving unique solution (61).

For (iv), the subsets of $\{1,\ldots,n\}$ with ordered type β have the form

$$S \oplus q = \{\beta_1 + q, \beta_1 + \beta_2 + q, \ldots, \beta_1 + \cdots + \beta_k + q\} \pmod{n} \tag{64}$$

where $S = \{1+\beta_1, 1+\beta_1+\beta_2, \ldots\}$; every sum is reduced modulo n to a number between 1 and n; and q ranges over all integers. Values $q = 0, 1, 2, \ldots, \frac{n}{\ell(\beta)/\operatorname{per}(\beta)} - 1$ give distinct sets $S \oplus q$ and all other values of q repeat one of these. So $\mathring{B}_\beta^{(g)} = \sum_q \mathring{B}_{n,S\oplus q}$ over these values of q, giving (iv).

For (v), $\mathring{b}_\mu^{(g)} = \sum_S \mathring{b}_{n,S}^{(g)}$ where S ranges over subsets of $\{1,\ldots,n\}$ of unordered circular type μ. All of these have $k = |S|$ breakpoints, so $\mathring{b}_\mu^{(g)}$ is $\mathring{b}_{k,k}^{(g)}$ times the number, $\mathring{M}(\mu)$, of such sets. The formula for $\mathring{M}(\mu)$ is proven below in Proposition 9.5. □

Proposition 9.5. *For $n > 0$, the number of subsets of $\{1,\ldots,n\}$ with unordered circular type $\mu \in \mathcal{P}_n$ is $\mathring{M}(\mu) = \frac{n}{\ell(\mu)} M(\mu)$.*

Proof. Let $k = \ell(\mu)$. Let E be the subset of $\mathring{\mathcal{C}}_n$ consisting of compositions that are permutations of μ. Each $\beta \in E$ has $\operatorname{per}(\beta)$ distinct cyclic shifts, so the total number of compositions in $\mathcal{C}_n$ obtained by permuting parts of μ is $M(\mu) = \sum_{\beta \in E} \operatorname{per}(\beta)$.

Circular arrangements of unordered type μ have ordered type β for some $\beta \in E$. There are $\frac{n}{\ell(\beta)/\operatorname{per}(\beta)}$ distinct breakpoint sets with circular ordered type β, shown in (64). Note that $\ell(\beta) = \ell(\mu)$ for all $\beta \in E$. So

$$\mathring{M}(\mu) = \sum_{\beta \in E} \frac{n}{\ell(\beta)/\operatorname{per}(\beta)} = \sum_{\beta \in E} \frac{n}{\ell(\mu)/\operatorname{per}(\beta)} = \frac{n}{\ell(\mu)} \sum_{\beta \in E} \operatorname{per}(\beta) = \frac{n}{\ell(\mu)} M(\mu) \ . \tag{65}$$

□

9.2 Signed circular arrangements by ordered type

For every n, the circular identity arrangement consists of one circular strip (Note 9.2), while all other arrangements consist of two or more linear strips. We introduce new variables with ring accents for weights of circular strips.

For weights of signed arrangements, let $\vec{v} = (v_1, v_2, v_3, \ldots)$ and $\mathring{\vec{v}} = (\mathring{v}_0, \mathring{v}_1, \mathring{v}_2, \ldots)$ be infinite sequences of commuting indeterminates. Let $\vec{V} = (V_1, V_2, V_3, \ldots)$ and $\mathring{\vec{V}} = (\mathring{V}_0, \mathring{V}_1, \mathring{V}_2, \ldots)$ be infinite sequences of noncommuting indeterminates. For convenience, set $v_0 = V_0 = 1$ (note that $\mathring{v}_0$ and $\mathring{V}_0$ are still indeterminates, not 1). Define monomials

V_β, v_μ as before ($V_\beta = V_{\beta_1}V_{\beta_2}\cdots$). Define $\vec{u}$, $\mathring{\vec{u}}$, $\vec{U}$, $\mathring{\vec{U}}$, U_α, u_λ similarly for unsigned arrangements.

In the noncommutative case, we define the *straightening* of a monomial V_γ to be $[V_\gamma]_\circ = V_\beta$, where $\gamma \in \mathcal{C}_{n,k}$ and $\beta \in \mathring{\mathcal{C}}_{n,k}$ is the lexicographically largest cyclic shift of γ. This corresponds to converting an ordered 1-type to an ordered type. For a formal series

$$f = \sum_{n=0}^{\infty} \mathring{k}_n \mathring{V}_n + \sum_{n=0}^{\infty}\sum_{\gamma\in\mathcal{C}_n} k_\gamma V_\gamma \quad \text{the straightening is} \quad [f]_\circ = \sum_{n=0}^{\infty} \mathring{k}_n \mathring{V}_n + \sum_{n=0}^{\infty}\sum_{\gamma\in\mathcal{C}_n} k_\gamma [V_\gamma]_\circ$$

where the coefficients (k's) are free of V's. For example, $[(5GV_{(2,1)} - V_3)(V_5 - V_{(4,1)})]_\circ = [5GV_{(2,1,5)} - 5GV_{(2,1,4,1)} - V_{(3,5)} + V_{(3,4,1)}]_\circ = 5GV_{(5,2,1)} - 5GV_{(4,1,2,1)} - V_{(5,3)} + V_{(4,1,3)}$. Define straightening for monomials and series in U_α's analogously. Straightening is not relevant in the commutative case.

Note that straightening commutes with sums, $[f+h]_\circ = [f]_\circ + [h]_\circ$, but does not commute with products: usually, $[fh]_\circ \neq [f]_\circ[h]_\circ$, so straightening must be done after computing all products. Multiplication has a cyclic property $[f_1 f_2 \cdots f_k]_\circ = [f_2 \cdots f_k f_1]_\circ$.

Let $\vec{\sigma} \in \mathring{\mathcal{B}}^{(g)}$. If $\vec{\sigma} = \mathrm{id}_n^{(g)}$ then its *ordered weight* is $\mathring{\omega}_B(\sigma) = \mathring{V}_n$ and its *unordered weight* is $\mathring{v}_n$. For $\vec{\sigma} \neq \mathrm{id}_n^{(g)}$, if $\vec{\sigma}$ has ordered 1-type γ, ordered type β, and unordered type μ, then its *ordered weight* is $\mathring{\omega}_B(\vec{\sigma}) = [V_\gamma]_\circ = V_\beta$ and its *unordered weight* is $\mathring{\omega}_b(\vec{\sigma}) = v_\mu$. Weights of sets of arrangements are defined analogously to the linear case. The generating function for the number of signed circular arrangements by ordered type is

$$\mathring{B}^{(g)}(\vec{V};\mathring{\vec{V}};t) = \sum_{n=0}^{\infty} t^n \mathring{B}_n^{(g)}(\vec{V};\mathring{\vec{V}}) = \sum_{n=0}^{\infty} \mathring{V}_n t^n + \sum_{n=1}^{\infty} t^n \sum_{\beta\in\mathring{\mathcal{C}}_n,\ \ell(\beta)>1} \mathring{B}_\beta^{(g)} V_{\beta_1}V_{\beta_2}\dots V_{\beta_{\ell(\beta)}} . \quad (66)$$

Since we summed over $\beta \in \mathring{\mathcal{C}}_n$ rather than $\mathcal{C}_n$, the monomials V_β are already straightened. We will give the analogous formula for the unordered case in Section 9.4.

Set $V(t) = \sum_{n=1}^\infty t^n V_n$, $\quad V'(t) = \sum_{n=1}^\infty n t^{n-1} V_n$, $\quad$ and $\mathring{V}(t) = \sum_{n=0}^\infty t^n \mathring{V}_n$.

The following theorem gives a formula for (66). Later we will apply ϕ^{-1} to it to derive a generating function for the unsigned case.

Theorem 9.6. *The generating function for the number of signed circular arrangements by ordered type is*

$$\begin{aligned}\mathring{B}^{(g)}(\vec{V};\mathring{\vec{V}};t) &= \left[\mathring{V}(t) + t\,V'(t)\left(-\bigl(1+V(t)\bigr)^{-1} + \sum_{r=1}^{\infty}(2^{r-1}(r-1)!)^{g-1}\frac{V(t)^{r-1}}{\bigl(1+V(t)\bigr)^{r+1}}\right)\right]_\circ \\ &= \left[\mathring{V}(t) + t\,V'(t)\left(-\bigl(1+V(t)\bigr)^{-1} + \frac{B^{(g)}(\vec{V};t)}{\bigl(1+V(t)\bigr)^{2}}\right)\right]_\circ \qquad (67)\end{aligned}$$

Proof. In (66), consider $\mathring{B}_\beta^{(g)} V_{\beta_1}V_{\beta_2}\dots V_{\beta_k}$, where $\beta \in \mathring{\mathcal{C}}_{n,k}$ with $k > 1$. By Corollary 9.4(iv), $\mathring{B}_\beta^{(g)} = \frac{n}{\ell(\beta)/\operatorname{per}(\beta)}\mathring{b}_{k,k}^{(g)}$. Let $d = \operatorname{per}(\beta)$, so β has k/d periods. A single period of β has sum $\beta_1 + \cdots + \beta_d = \frac{n}{k/d} = \frac{n}{\ell(\beta)/\operatorname{per}(\beta)}$, so $\mathring{B}_\beta^{(g)} = (\beta_1 + \cdots + \beta_d)\mathring{b}_{k,k}^{(g)}$. Then

$$\mathring{B}_\beta^{(g)} V_{\beta_1}V_{\beta_2}\dots V_{\beta_k} = \mathring{B}_\beta^{(g)} V_\beta = \textstyle\sum_\gamma \mathring{b}_{k,k}^{(g)}\,\gamma_1\,[V_\gamma]_\circ$$

where γ runs over $\{(\beta_i, \beta_{i+1}, \ldots, \beta_k, \beta_1, \cdots, \beta_{i-1}) :\ i = 1, \ldots, d\}$, the d distinct compositions that straighten to β. (Note that in that sum, γ_1 takes on values $\beta_1, \ldots, \beta_d$.) By (66),

$$\mathring{B}^{(g)}(\vec{V}; \mathring{\vec{V}}; t) = \sum_{n=0}^{\infty} \mathring{V}_n t^n + \sum_{n=1}^{\infty} t^n \sum_{\beta \in \mathring{\mathcal{C}}_n,\, \ell(\beta)>1} \mathring{B}^{(g)}_{\beta} V_\beta = \mathring{V}(t) + \sum_{k=2}^{\infty} \mathring{b}^{(g)}_{k,k} \sum_{n=1}^{\infty} t^n \sum_{\gamma \in \mathcal{C}_{n,k}} \gamma_1 [V_\gamma]_\circ$$

Note that $\mathring{b}^{(g)}_{1,1} = 0$, so we may extend this sum to $k = 1$. Also we commute straightening with the linear combination.

$$= \mathring{V}(t) + \left[\sum_{k=1}^{\infty} \mathring{b}^{(g)}_{k,k} \sum_{n=1}^{\infty} t^n \sum_{\gamma \in \mathcal{C}_{n,k}} \gamma_1 V_\gamma \right]_\circ = \mathring{V}(t) + \left[\sum_{k=1}^{\infty} \mathring{b}^{(g)}_{k,k} t\, V'(t) V(t)^{k-1} \right]_\circ$$

$$= \mathring{V}(t) + \left[t\, V'(t) \sum_{k=1}^{\infty} \mathring{b}^{(g)}_{k,k} V(t)^{k-1} \right]_\circ \tag{68}$$

Use (61) to evaluate the sum within that expression:

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathring{b}^{(g)}_{k,k} V(t)^{k-1} &= \sum_{k=1}^{\infty} (-1)^k V(t)^{k-1} + \sum_{k=1}^{\infty} \sum_{r=1}^{k} (-1)^{k-r} \binom{k}{r} (2^{r-1}(r-1)!)^{g-1} V(t)^{k-1} \\
&= -\big(1+V(t)\big)^{-1} + \sum_{r=1}^{\infty} (2^{r-1}(r-1)!)^{g-1} \sum_{k=r}^{\infty} (-1)^{k-r} \binom{k}{r} V(t)^{k-1} \\
&= -\big(1+V(t)\big)^{-1} + \sum_{r=1}^{\infty} (2^{r-1}(r-1)!)^{g-1} \frac{V(t)^{r-1}}{\big(1+V(t)\big)^{r+1}} \\
&= -\big(1+V(t)\big)^{-1} + \frac{B^{(g)}(\vec{V}; t)}{\big(1+V(t)\big)^2} \qquad \text{by (12).}
\end{aligned} \tag{69}$$

Plugging (69) into (68) gives (67). ☐

Specialization of (67) for the number of signed circular arrangements by size and number of strips will be considered in Section 9.5.

9.3 Unsigned circular arrangements by ordered type

In this section, we will extend Theorem 5.2 and Corollary 5.3 to compute the ordered weight generating function for unsigned circular arrangements.

For weights of unsigned circular arrangements, notation is defined analogously to the signed case but with U's instead of V's and a's instead of b's (with the same conventions on upper/lowercase, fonts, boldface, ring accents, etc.). $\mathring{A}^{(g)}_{\alpha}$ is the number of unsigned circular arrangements of ordered type α. This has generating function

$$\mathring{A}^{(g)}(\vec{U}; \mathring{\vec{U}}; t) = \sum_{n=0}^{\infty} t^n \mathring{A}^{(g)}_n(\vec{U}; \mathring{\vec{U}}) = \sum_{n=0}^{\infty} \mathring{U}_n t^n + \sum_{n=1}^{\infty} t^n \sum_{\alpha \in \mathring{\mathcal{C}}_n,\, \ell(\alpha)>1} \mathring{A}^{(g)}_{\alpha} U_{\alpha_1} U_{\alpha_2} \ldots U_{\alpha_{\ell(\alpha)}} \tag{70}$$

We will derive a formula for this by applying ϕ^{-1} to the signed weight formula (67). To do this, we must first compute $\phi(\mathring{U}_n)$.

Theorem 9.7. *(i) Consider the unsigned circular identity (n,g)-arrangement $\mathrm{id}_n^{(g)}$. Its ordered weight is $\mathring{U}_n$, and the ordered weight of all signages of it is $\phi(\mathring{U}_n) = \mathring{V}_n + W_n$ where $W_0 = W_1 = 0$, $W_2 = GV_1V_1$, and for $n > 2$,*

$$W_n = \sum_{k=1}^{n} \sum_{\gamma \in \mathcal{C}_{n,k}} G^k \gamma_1 \left[V_{\gamma_1 - 1} V_1 V_{\gamma_2 - 1} V_1 \cdots V_{\gamma_k - 1} V_1\right]_\circ \tag{71}$$

$$= \sum_{k=1}^{n} \sum_{\beta \in \mathring{\mathcal{C}}_{n,k}} \frac{n}{k/\operatorname{per}(\beta)} G^k V_{\beta_1 - 1} V_1 V_{\beta_2 - 1} V_1 \cdots V_{\beta_k - 1} V_1 \tag{72}$$

(ii) Let $\vec{\pi} \neq \mathrm{id}_n^{(g)}$ be an unsigned circular (n,g)-arrangement with ordered type α. The ordered weight of all signages of $\vec{\pi}$ is $[\phi(U_{\alpha_1})\phi(U_{\alpha_2})\cdots]_\circ$, where $\phi(U_i)$ is given by (20).

(iii) $[\phi\big(\mathring{A}^{(g)}(\vec{U}; \mathring{\vec{U}}; t)\big)]_\circ = \mathring{B}^{(g)}(\vec{V}; \mathring{\vec{V}}; t)$.

Proof. Parts (ii) and (iii) have similar proofs to the linear case (Theorem 5.2(ii) and Corollary 5.3), except that we straighten the weights to account for circularity. Part (i), however, is new.

When $n = 0$ or 1, the only signage is $\mathrm{id}_n^{(g)}$, so the total weight of signages is $\phi(\mathring{U}_n) = \mathring{V}_n$.

When $n = 2$, all $2^{g-1} = G + 1$ elements of $\mathring{B}_2^{(g)}$ are signages of the unsigned identity. One of them is the signed identity, with weight $\mathring{V}_2$. The remaining G of them each have weight V_1V_1, for a total weight $\phi(\mathring{U}_2) = \mathring{V}_2 + GV_1V_1$.

The rest of the proof concerns $n \geq 3$. Signages of the unsigned circular $\mathrm{id}_n^{(g)}$ can be described by sign vectors $\vec{\epsilon}_j = (\epsilon_{1j}, \ldots, \epsilon_{gj})$ for $j = 1, \ldots, n$, where $\epsilon_{1j} = 1$.

Consider any signage $\vec{\sigma}$ of $\mathrm{id}_n^{(g)}$. Let $j_1 < \cdots < j_k$ be the entries (between 1 and n inclusive) for which $\vec{\epsilon}_j \neq \vec{\epsilon}_+$. If $k = 0$ then all $\vec{\epsilon}_j = \vec{\epsilon}_+$ and we have the signed circular identity (n,g)-arrangement $\vec{\sigma} = \mathrm{id}_n^{(g)}$, with weight $\mathring{V}_n$.

Otherwise, let $\gamma = (j_1 + n - j_k, j_2 - j_1, j_3 - j_2, \ldots, j_k - j_{k-1})$. Then $\vec{\sigma}$ has 1-type $\rho = (\gamma_1 - 1, 1, \gamma_2 - 1, 1, \ldots, \gamma_{k-1} - 1, 1, \gamma_k - 1)$ (except that we omit any 0's that arise from $\gamma_r - 1$ with $\gamma_r = 1$) and weight $[V_\rho]_\circ = \left[V_{\gamma_1 - 1} V_1 V_{\gamma_2 - 1} V_1 \cdots V_{\gamma_k - 1}\right]_\circ$ (where the 0's in ρ are effectively removed in the product on the right by having set $V_0 = 1$). If we view the arrangement shown at the top of Example 5.1 as circular, then $S = \{3, 4, 8\}$, $\gamma = (4, 1, 4)$, $\rho = (3, 1, 1, 3, 1)$ (omitting 0's in $(3, 1, 0, 1, 3, 1)$), the ordered type is $(3, 1, 3, 1, 1)$, and the weight is $V_3V_1V_0V_1V_3V_1 = V_3V_1V_3V_1V_1 = [V_\rho]_\circ$.

The total number of signages with signs $\neq \vec{\epsilon}_+$ precisely on entries $S = \{j_1, \ldots, j_k\}$ is G^k. The sets S' leading to same γ as S are $S \oplus q = \{j_1 + q, \ldots, j_k + q\}$ where $-(j_1 - 1) \leq q \leq n - j_k$; there are $\gamma_1 = n - j_k + j_1$ of these. So the weight of all signages with the same value of γ is $\gamma_1 \left[V_{\gamma_1 - 1} V_1 V_{\gamma_2 - 1} V_1 \cdots V_{\gamma_k - 1}\right]_\circ$.

Eq. (71) sums this over all possibilities of at least one non-$\vec{\epsilon}_+$ sign to get the total weight W_n of all such signages.

If we use ordered type instead of ordered 1-type of S, there would be $\frac{n}{\ell(\beta)/\operatorname{per}(\beta)}$ distinct shifts of S with the same ordered type β, giving sum (72) instead. $\square$

Computing $\mathring{A}^{(g)}(\vec{U};\vec{\mathring{U}};t) = [\phi^{-1}\big(\mathring{B}^{(g)}(\vec{V},\vec{\mathring{V}};t)\big)]_\circ$ leads to the following result:

Theorem 9.8. *A generating function for the number of unsigned circular arrangements by ordered type is*

$$\mathring{A}^{(g)}(\vec{U};\vec{\mathring{U}};t) = \Big[\big((1+U(t))^{-1}tU'(t)(1+U(t))^{-1} + \widetilde{G}U_1t\big)\big(A^{(g)}(\vec{U};t) - 1 - U(t)\big) + (U(t)-U_1t)\widetilde{G}U_1t\Big]_\circ + \mathring{U}(t) \tag{73}$$

The proof is tedious, so we postpone it to Appendix A.3. Given this theorem, for any n we may compute $\mathring{A}^{(g)}_\alpha$ for all $\alpha \in \mathring{\mathcal{C}}_n$ by a procedure similar to the one in Section 7:

1. Compute $\mathring{B}^{(g)}_n(\vec{V};\vec{\mathring{V}}) = \omega_B(\mathring{\mathcal{B}}^{(g)}_n) = \mathring{V}_n + \sum_{\beta\in\mathring{\mathcal{C}}_n,\ell(\beta)>1} \mathring{B}^{(g)}_\beta V_{\beta_1}V_{\beta_2}\dots V_{\beta_{\ell(\beta)}}$.
2. Compute $\mathring{A}^{(g)}_n(\vec{U};\vec{\mathring{U}}) = [\phi^{-1}\big(\mathring{B}^{(g)}_n(\vec{V};\vec{\mathring{V}})\big)]_\circ$ and collect terms by monomials in the U's and $\mathring{U}_n$. The coefficient of U_α is $\mathring{A}_\alpha$, and the coefficient of $\mathring{U}_n$ is $\mathring{A}_{C(n)} = 1$.

9.4 Circular arrangements by unordered type

In this section, we give generating functions for the number of signed or unsigned circular arrangements by unordered type. The results of the preceding sections go through to unordered types by letting the variables commute. We use commuting lowercase variables $u_n, \mathring{u}_n, u(t), \mathring{u}(t)$ in the unsigned case (and v's for the signed case), as previously defined.

The *commutative specialization* is to specialize $U_n \to u_n$, $\mathring{U}_n \to \mathring{u}_n$, $V_n \to v_n$, and $\mathring{V}_n \to \mathring{v}_n$. Straightening is not needed in the commutative case, so we simply drop the straightening operations.

A generating function for the number of signed circular arrangements by unordered type is obtained by specializing (66) and Theorem 9.6 to commutative variables:

$$\mathring{b}^{(g)}(\vec{v};\vec{\mathring{v}};t) = \sum_{n=0}^{\infty} t^n \mathring{b}^{(g)}_n(\vec{v};\vec{\mathring{v}}) = \sum_{n=0}^{\infty} \mathring{v}_n t^n + \sum_{n=1}^{\infty} t^n \sum_{\mu\in\mathcal{P}_n,\,\ell(\mu)>1} \mathring{b}^{(g)}_\mu v_{\mu_1}v_{\mu_2}\dots v_{\mu_{\ell(\mu)}} \tag{74}$$

$$= \mathring{v}(t) + t\,v'(t)\left(-(1+v(t))^{-1} + \frac{b^{(g)}(\vec{v};t)}{(1+v(t))^2}\right) . \tag{75}$$

A generating function for unsigned circular arrangements by unordered type is obtained by specializing (70) and Theorem 9.8 to commutative variables:

$$\mathring{a}^{(g)}(\vec{u};\vec{\mathring{u}};t) = \sum_{n=0}^{\infty} t^n \mathring{a}^{(g)}_n(\vec{u};\vec{\mathring{u}}) = \sum_{n=0}^{\infty} \mathring{u}_n t^n + \sum_{n=1}^{\infty} t^n \sum_{\lambda\in\mathcal{P}_n,\,\ell(\lambda)>1} \mathring{a}^{(g)}_\lambda u_{\lambda_1}u_{\lambda_2}\dots u_{\lambda_{\ell(\lambda)}} \tag{76}$$

$$= \mathring{u}(t) + \left(\frac{tu'(t)}{(1+u(t))^2} + \widetilde{G}u_1t\right)\big(a^{(g)}(\vec{u};t) - 1 - u(t)\big) + (u(t)-u_1t)\widetilde{G}u_1t \tag{77}$$

The ring homomorphism ϕ extends to the commutative case for circular arrangements:

Theorem 9.9. *(i) Consider the unsigned circular identity (n,g)-arrangement $\mathrm{id}_n^{(g)}$. Its unordered weight is $\mathring{u}_n$. The unordered weight of all signages of it is $\phi(\mathring{u}_n) = \mathring{v}_n + w_n$ where $w_0 = w_1 = 0$, $w_2 = Gv_1v_1$, and for $n > 2$,*

$$w_n = \sum_{k=1}^{n} \sum_{\mu \in \mathcal{P}_{n,k}} G^k \mathring{M}(\mu) v_{\mu_1 - 1} v_{\mu_2 - 1} \cdots v_{\mu_k - 1} {v_1}^k \tag{78}$$

(ii) Let $\vec{\sigma} \neq \mathrm{id}_n^{(g)}$ be an unsigned circular (n,g)-arrangement with unordered type λ. The unordered weight of all signages of $\vec{\sigma}$ is $\phi(u_{\lambda_1})\phi(u_{\lambda_2})\cdots$, where $\phi(u_i)$ is given by (47).

(iii) $\phi\big(\mathring{a}^{(g)}(\vec{u};\mathring{\vec{u}};t)\big) = \mathring{b}^{(g)}(\vec{v};\mathring{\vec{v}};t)$.

Proof. This is a specialization of Theorem 9.7 to commutative variables.

Part (i) requires additional proof. Consider the commutative specialization of (72). For a partition $\mu \in \mathcal{P}_{n,k}$, let E be the set of $\beta \in \mathring{\mathcal{C}}_{n,k}$ that are permutations of the parts of μ. The terms of (72) with $\beta \in E$ are collected into a single term when the variables become commutative; by (65), $\sum_{\beta \in E} \frac{n}{k/\operatorname{per}(\beta)} = \mathring{M}(\mu)$, giving (78). □

Finally, $\mathring{a}_\lambda^{(g)}$ can be computed by a procedure similar to the one at the end of Section 9.3: let $b_n(\vec{v},\mathring{\vec{v}}) = \mathring{v}_n + \sum_{\mu \in \mathring{\mathcal{C}}_n, \ell(\mu)>1} \mathring{b}_\mu^{(g)} v_\mu$, compute $\phi^{-1}\big(\mathring{b}_n^{(g)}(\vec{v};\mathring{\vec{v}})\big)$, and collect terms by monomials in the u's. The coefficient of u_λ is $a_\lambda^{(g)}$ and the coefficient of $\mathring{u}_n$ is $a_{C(n)}^{(g)} = 1$.

9.5 Number of strips in circular arrangements

We may specialize the equations for $\mathring{a}^{(g)}(\vec{u};\mathring{\vec{u}};t)$ and $\mathring{b}^{(g)}(\vec{v};\mathring{\vec{v}};t)$ to get generating functions for the number of strips and number of incompressible arrangements.

For the number of (linear) strips in $\mathring{\mathcal{B}}$, specialize $\mathring{v}_n \to z^0$ $(n \geq 0)$ and $v_n \to z$ $(n \geq 1)$ to get $\mathring{v}(t) \to 1/(1-t)$; $v(t) \to zt/(1-t)$; and $v'(t) \to z/(1-t)^2$. Apply this to (75):

$$\begin{aligned}
\mathring{b}^{(g)}(t,z) &= \sum_{n=0}^{\infty} \mathring{b}_n^{(g)}(z)\, t^n = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \mathring{b}_{n,k}^{(g)}\, z^k t^n \\
&= \frac{1}{1-t(1-z)} + \frac{zt}{(1-t(1-z))^2} \sum_{r=1}^{\infty} (2^{r-1}(r-1)!)^{g-1} \frac{(zt)^{r-1}}{(1-t(1-z))^{r-1}} \\
&= \frac{1}{1-t(1-z)} + \frac{zt}{(1-t(1-z))^2} b^{(g)}(t,z)\ .
\end{aligned} \tag{79}$$

Take the coefficient of t^n in the Maclaurin expansion to get

$$\mathring{b}_n^{(g)}(z) = (1-z)^n + \sum_{r=0}^{n-1} (2^r r!)^{g-1} \binom{n}{r+1} z^{r+1} (1-z)^{n-r-1}\ . \tag{80}$$

For incompressible arrangements, we specialize $\mathring{v}_n \to 0$ for $n \geq 0$, $v_1 \to 1$, $v_n \to 0$ for $n > 1$. This gives $\mathring{v}(t) \to 0$, $v(t) \to t$, and $v'(t) \to 1$. Apply this to (75):

$$\begin{aligned}\mathring{\mathrm{IB}}^{(g)}(t) = \sum_{n=0}^{\infty} \mathring{b}^{(g)}_{n,n} t^n &= -\frac{t}{1+t} + t\sum_{r=1}^{\infty}(2^{r-1}(r-1)!)^{g-1}\frac{t^{r-1}}{(1+t)^{r+1}} \\ &= -\frac{t}{1+t} + \frac{t}{(1+t)^2}\ \mathrm{IB}^{(g)}(t)\end{aligned}$$

where $\mathrm{IB}^{(g)}(t)$ is given by (15). This rearranges to $\mathrm{IB}^{(g)}(t) = 1 + t + \frac{(1+t)^2}{t}\,\mathring{\mathrm{IB}}(t)$. Taking the coefficient of t^n relates incompressible signed linear and circular arrangements:

$$b^{(g)}_{n,n} = \mathring{b}^{(g)}_{n+1,n+1} + 2\mathring{b}^{(g)}_{n,n} + \mathring{b}^{(g)}_{n-1,n-1} \qquad \text{(for } n \geq 1). \tag{81}$$

We make similar specializations for unsigned circular arrangements. A generating function for $\mathring{a}^{(g)}_{n,k}$ is obtained by making these specializations in (77): $\mathring{u}_n \to z^0$ ($n \geq 0$) and $i_n \to z$ ($n \geq 1$), which give $\mathring{u}(t) \to 1/(1-t)$ and $u(t) \to zt/(1-t)$. Plugging these into (77) and simplifying gives

$$\begin{aligned}\mathring{a}^{(g)}(t,z) &= \sum_{n=0}^{\infty} \mathring{a}^{(g)}_n(z)\, t^n = \sum_{n=0}^{\infty}\sum_{k=0}^{n} \mathring{a}^{(g)}_{n,k} z^k t^n \\ &= \frac{1}{1-t(1-z)} - \widetilde{G}zt(1+zt) \\ &\quad + \left(\widetilde{G}zt + \frac{zt}{(1-t(1-z))^2}\right)\sum_{r=0}^{\infty} r!^{g-1}\left(\frac{zt(1+Gt(1-z))}{1-t(1-z)}\right)^r \\ &= \frac{1}{1-t(1-z)} - \widetilde{G}zt(1+zt) + \left(\widetilde{G}zt + \frac{zt}{(1-t(1-z))^2}\right) a^{(g)}(t,z)\ .\end{aligned} \tag{82}$$

Expanding this as a Maclaurin series in t and taking the coefficient of t^n gives $\mathring{a}^{(g)}_n(z) = 1$ for $n = 0, 1, 2$, and the following for $n \geq 3$:

$$\mathring{a}^{(g)}_n(z) = (1-z)^n + \sum_{r=0}^{n-1} r!^{g-1} z^{r+1}(1-z)^{n-r-1} \sum_{i=0}^{\min(r,n-r-1)} G^i \tbinom{r}{i}\left(\widetilde{G}\tbinom{n-i-2}{r-1} + \tbinom{n-i}{r+1}\right) \tag{83}$$

Expanding that as a polynomial in z and taking the coefficient of z^k gives $\mathring{a}^{(g)}_{n,0} = 1$ (all $n \geq 0$), $\mathring{a}^{(g)}_{n,1} = 0$ (all $n \geq 0$), $\mathring{a}^{(g)}_{2,k} = 0$ (all $k \geq 1$), and for $n \geq 3$, $k \geq 0$,

$$\mathring{a}^{(g)}_{n,k} = (-1)^k\tbinom{n}{k} + \sum_{r=0}^{k-1} r!^{g-1}(-1)^{k-r-1}\tbinom{n-r-1}{k-r-1} \sum_{i=0}^{\min(r,n-r-1)} G^i\tbinom{r}{i}\left(\widetilde{G}\tbinom{n-i-2}{r-1} + \tbinom{n-i}{r+1}\right) \tag{84}$$

The specialization for the number of incompressible unsigned circular arrangements is $\mathring{u}_n \to 0$ ($n \geq 0$), $u_1 \to 1$ and $u_n \to 0$ ($n > 1$). This gives $\mathring{u}(t) \to 0$ and $u(t) \to t$. Making these specializations in (77) and simplifying gives a generating function for incompressible unsigned circular arrangements (see (34) for the linear case, $\mathrm{IA}^{(g)}(t)$):

$$\mathrm{I\mathring{A}}^{(g)}(t) = \sum_{n=0}^{\infty} \mathring{a}_{n,n}^{(g)} t^n = t\Big(\frac{1}{(1+t)^2} + \widetilde{G}\Big)\Big(\mathrm{IA}^{(g)}(t) - (1+t)\Big) \tag{85}$$

10 Expected number of strips

We will compute the expected number of linear strips in all four types of arrangements studied here, under the uniform distribution on all arrangements.

Each of $a_n^{(g)}(z), b_n^{(g)}(z), \mathring{a}_n^{(g)}(z), \mathring{b}_n^{(g)}(z)$ has the form $f_n(z) = \sum_{k=0}^{n} f_{n,k}^{(g)} z^k$, where $f_{n,k}^{(g)}$ is a count ≥ 0. Under the uniform distribution, the corresponding probability generating function is $F_n(z) = f_n(z)/f_n(1)$. The expected number of strips is $\mu = F_n'(1) = f_n'(1)/f_n(1)$ and the variance is $\sigma^2 = F_n''(1) + \mu - \mu^2 = f_n''(1)/f_n(1) + \mu - \mu^2$.

For signed linear arrangements, $\mathcal{B}_n^{(g)}$, use $f_n(z) = b_n^{(g)}(z)$ as given in (14). Then $b_n^{(g)}(1) = (2^n n!)^{g-1}$, so

$$F_n(z) = b_n^{(g)}(z)/(2^n n!)^{g-1} = \sum_{r=1}^{n} \frac{(2^r r!)^{g-1}}{(2^n n!)^{g-1}} \binom{n-1}{r-1} z^r (1-z)^{n-r}$$

On differentiating and plugging in $z = 1$, most terms vanish due to the powers of $1 - z$:

$$\mu = F_n'(1) = n - \frac{n-1}{(2n)^{g-1}}$$

$$F_n''(1) = n(n-1) - \frac{2(n-1)^2}{(2n)^{g-1}} + \frac{(n-1)(n-2)}{(4n(n-1))^{g-1}}$$

$$\sigma^2 = F_n''(1) + F_n'(1) - (F_n'(1))^2 = -\frac{(n-1)^2}{(2n)^{2(g-1)}} + \frac{n-1}{(2n)^{g-1}} + \frac{(n-1)(n-2)}{(4n(n-1))^{g-1}}$$

The formulas hold for $n \geq 2$. For $n = 0$, $(\mu, \sigma^2) = (0, 0)$, and for $n = 1$, $(\mu, \sigma^2) = (1, 0)$.

The above computations also could have been done in generating function form for all n simultaneously from (13).

For unsigned linear arrangements, $\mathcal{A}_n^{(g)}$, use $f_n(z) = a_n^{(g)}(z)$ as given in (36), and $f_n(1) = n!^{(g-1)}$. The computations are analogous to those above (but messier) and result in $(\mu, \sigma^2) = (0, 0)$ for $n = 0$; $(\mu, \sigma^2) = (1, 0)$ for $n = 1$; and the following for $n \geq 2$:

$$\mu = n - \frac{2^{g-1}(n-1)}{n^{g-1}} \tag{86}$$

$$\sigma^2 = \frac{2^{g-1}(n-1)}{n^{g-1}} - \frac{4^{g-1}(n-1)}{n^{2(g-1)}} + \frac{2^{g-1}(n-2)(2^{g-1}(n-3)+2)}{(n(n-1))^{g-1}} \tag{87}$$

Note that Wolfowitz, 1942 [33, Eqs. (7.9,7.10)] obtained the $g = 2$ case of (86)–(87).

For signed circular arrangements, $\mathring{\mathcal{B}}_n^{(g)}$, use $f_n(z) = \mathring{b}_n^{(g)}(z)$ (Eq. (80)) and $f_n(1) = (2^{n-1}(n-1)!)^{g-1}$ for $n \geq 1$ (note $f_0(1) = 1$). Then for $n = 0, 1$ we have $(\mu, \sigma^2) = (0, 0)$; for $n = 2$ we have $(\mu, \sigma^2) = (2 - 2^{2-g}, 2^{3-g} - 4^{2-2g})$; and for $n \geq 3$,

$$\mu = n\left(1 - \frac{1}{2^{g-1}(n-1)^{g-1}}\right)$$

$$\sigma^2 = \frac{n}{2^{g-1}(n-1)^{g-1}} + \frac{n(n-1)}{4^{g-1}(n-1)^{g-1}(n-2)^{g-1}} - \frac{n^2}{4^{g-1}(n-1)^{2(g-1)}}$$

For unsigned circular arrangements, $\mathring{\mathcal{A}}_n^{(g)}$, use $f_n(z) = \mathring{a}_n^{(g)}(z)$ as given in (83), and $f_n(1) = ((n-1)!/2)^{g-1}$ for $n \geq 3$ (note $f_n(1) = 1$ for $n = 0, 1, 2, 3$). Then for $n = 0, 1, 2, 3$, we have $(\mu, \sigma^2) = (0, 0)$; and for $n \geq 3$,

$$\mu = n\left(1 - \frac{2^{g-1}}{(n-1)^{g-1}}\right)$$

$$\sigma^2 = \frac{n2^{g-1}}{(n-1)^{g-1}} - \frac{n^2 4^{g-1}}{(n-1)^{2(g-1)}} + \frac{n\big((n-3)4^{g-1} + 2(2^{g-1})\big)}{(n-1)^{g-1}(n-2)^{g-1}}$$

The results above for (un)signed circular arrangements are for the expected number of *linear* strips; the circular identity counts as one *circular* strip and zero linear strips.

11 Recursions and differential equations

For fixed $g \geq 2$, we will describe how to construct both a recursion and a mixed differential/recursion equation for each of $a_n^{(g)}(z), b_n^{(g)}(z), \mathring{a}_n^{(g)}(z), \mathring{b}_n^{(g)}(z)$ in terms of smaller n. Although we have given direct formulas for these, it is also useful to have recursion equations. We'll work out the case of $a_n^{(2)}(z)$ and sketch the details of the others. However, our Maple software computes all of them. We begin by rewriting (35) as

$$a^{(g)}(t,z) = \sum_{r=0}^{\infty} r!^{g-1} K(t,z)^r \qquad \text{where} \quad K(t,z) = \frac{zt(1+Gt(1-z))}{1-t(1-z)}\ . \tag{88}$$

First we will construct an ODE in t of order $g-1$ that $a^{(g)}(t,z)$ satisfies. Then we will turn that into a recursion in n that the $a_n^{(g)}(z)$'s satisfy.

Consider the operator P_K on functions of t and z:

$$P_K F(t,z) = \frac{1}{\partial K/\partial t}\frac{\partial}{\partial t}\big(K(t,z)F(t,z)\big) = \left(1 + \frac{K(t,z)}{\partial K/\partial t}\frac{\partial}{\partial t}\right) F(t,z)\ . \tag{89}$$

Since $\frac{\partial}{\partial t}(K(t,z)K(t,z)^r) = (r+1)K(t,z)^r \frac{\partial K}{\partial t}$, we have that $P_K(K(t,z)^r) = (r+1)K(t,z)^r$ and $(P_K)^{g-1}K(t,z)^r = (r+1)^{g-1}K(t,z)^r$. Thus, the operator

$$K(t,z){P_K}^{g-1} = K(t,z)\left(1 + \frac{K(t,z)}{\partial K/\partial t}\frac{\partial}{\partial t}\right)^{g-1}$$

promotes term r of the sum in (88) to term $r+1$. The only term of the sum not in the image of this operator is the $r=0$ term, which evaluates to $0!^{g-1}K(t,z)^0 = 1$. So

$$K(t,z)\left(1+\frac{K(t,z)}{\partial K/\partial t}\frac{\partial}{\partial t}\right)^{g-1} a^{(g)}(t,z)+1 = a^{(g)}(t,z)\ . \tag{90}$$

We will work out the $g=2$ case. Plug $g=2$ into (90) to obtain

$$K(t,z)\,a^{(g)}(t,z)+K(t,z)\,\frac{K(t,z)}{\partial K/\partial t}\,\frac{\partial a^{(g)}}{\partial t}(t,z)+1 = a^{(g)}(t,z)\ .$$

Plug $K(t,z)$ (Eq. (88)) and $G=2^{g-1}-1=1$ into this, clear denominators, and simplify:

$$\begin{aligned}\big(1+(1-2z)t+3(z-1)t^2-(z-1)^2(2z-1)t^3-z(z-1)^3t^4\big)a^{(g)}(t,z)&\\ +\big(-zt^2+z(z-1)t^3+z(z-1)^2t^4-z(z-1)^3t^5\big)\frac{\partial a^{(g)}}{\partial t}(t,z)&\\ =1-(z-1)t-3\,(z-1)^2t^2-(z-1)^3t^3\ .&\end{aligned} \tag{91}$$

Plug $a^{(g)}(t,z)=\sum_{n=0}^{\infty} t^n a_n^{(g)}(z)$ into (91) and use standard series methods to find a recurrence equation for $a_n^{(g)}(z)$: expand the sums, combine powers of t, and take coefficients of t^n. For convenience, set $a_n^{(g)}(z)=0$ when $n<0$. We obtain that for all integers n,

$$\begin{aligned}a_n^{(2)}(z)+(1-z-zn)a_{n-1}^{(2)}(z)-(1-z)\big((n-2)z+3\big)a_{n-2}^{(2)}(z)&\\ +(1-z)^2\big((n-5)z+1\big)a_{n-3}^{(2)}(z)+z(1-z)^3(n-3)a_{n-4}^{(2)}(z)&\\ =\delta_n-(z-1)\delta_{n-1}-3(z-1)^2\delta_{n-2}-(z-1)^3\delta_{n-3}&\end{aligned} \tag{92}$$

where $\delta_0=1$, $\delta_x=0$ for $x\neq 0$. So for $n\geq 4$, we have homogeneous recursion equations

$$\begin{aligned}a_n^{(2)}(z)+(1-z-zn)a_{n-1}^{(2)}(z)-(1-z)\big((n-2)z+3\big)a_{n-2}^{(2)}(z)&\\ +(1-z)^2\big((n-5)z+1\big)a_{n-3}^{(2)}(z)+z(1-z)^3(n-3)a_{n-4}^{(2)}(z)=0 \qquad (n\geq 4).&\end{aligned} \tag{93}$$

The inhomogeneities in (92) allow us to compute $a_0^{(2)}(z),\ldots,a_3^{(2)}(z)$. Plug in $n=0$:

$$\begin{aligned}a_0^{(2)}(z)+(1-z)a_{-1}^{(2)}(z)-(1-z)(-2z+3)a_{-2}^{(2)}(z)&\\ +(1-z)^2(-5z+1)a_{-3}^{(2)}(z)+z(1-z)^3(-3)a_{-4}^{(2)}(z)=1\ .&\end{aligned}$$

Since $a_{-1}^{(2)}(z)=\cdots=a_{-4}^{(2)}(z)=0$, this gives $a_0^{(2)}(z)=1$. Plugging in $n=1,2,3$ and using the corresponding inhomogeneities on the right gives initial conditions $a_0^{(2)}(z)=1$, $a_1^{(2)}(z)=z$, $a_2^{(2)}(z)=2z$, and $a_3^{(2)}(z)=2z+4z^2$.

For a fixed N, iterating this recursion gives a more efficient way to tabulate all $a_{n,k}^{(2)}$ with $0\leq k\leq n\leq N$ than repeated use of (37) or (36). However, those equations apply to general g, while a new recursion must be computed for each g. Analyzing this procedure further would show that for each g, we get a recursion of order $4(g-1)$ in n where the coefficient of $a^{(2)}_{n-i}(z)$ $(i = 0, \ldots, 4(g-1))$ has degree $\leq 4(g-1)$ in z and degree $\leq g-1$ in n. The recursion is homogeneous for $n \geq 4(g-1)$ and has initial conditions for $n < 4(g-1)$.

Note: For unsigned linear arrangements on $g = 2$ genomes, Eq. (93) is equivalent to a recurrence given by Riordan, 1965 [28] in different notation. Our method for deriving it is more general, as it may be applied to obtain recurrences for signed and unsigned, circular or linear, (n, g)-arrangements for all $g \geq 2$.

In the procedure illustrated above, we may use z-derivatives instead of t-derivatives throughout. That leads to this replacement for (90):

$$K(t,z)\left(1 + \frac{K(t,z)}{\partial K/\partial z}\,\frac{\partial}{\partial z}\right)^{g-1} a^{(g)}(t,z) + 1 = a^{(g)}(t,z)\ . \tag{94}$$

For $g = 2$, evaluating that and clearing denominators gives

$$\begin{aligned}\Big(1-(2z+1)t+(3z-1)t^2-(z-1)(2z^2-z+1)t^3-z(z-1)^3t^4\Big)a^{(2)}(t,z)\\ +\Big(-z^2t+z^2(z-1)t^2+z^2(z-1)^2t^3-z^2(z-1)^3t^4\Big)\frac{\partial a^{(2)}}{\partial z}(t,z)\\ = 1-(z+1)t-(3z-1)(z-1)t^2-(z-1)^3t^3\ . \end{aligned}\tag{95}$$

Plugging in $a^{(g)}(t,z) = \sum_{n=0}^{\infty} a^{(g)}_n(z)\, t^n$ leads to the mixed differential/recursion equation

$$\begin{aligned} a^{(2)}_n(z)-(2z+1)a^{(2)}_{n-1}(z)+(3z-1)a^{(2)}_{n-2}(z)-(z-1)(2z^2-z+1)a^{(2)}_{n-3}(z)-(z-1)^3 z a^{(2)}_{n-4}(z)\\ -z^2\frac{\partial a^{(2)}_{n-1}}{\partial z}+z^2(z-1)\frac{\partial a^{(2)}_{n-2}}{\partial z}+z^2(z-1)^2\frac{\partial a^{(2)}_{n-3}}{\partial z}-z^2(z-1)^3\frac{\partial a^{(2)}_{n-4}}{\partial z}\\ =\delta_n-(z+1)\delta_{n-1}-(3z-1)(z-1)\delta_{n-2}-(z-1)^3\delta_{n-3}\ . \end{aligned}\tag{96}$$

For fixed g, we may similarly derive both a recursion and a mixed differential/recursion equation for $b^{(g)}_n(z)$. By (13), $b^{(g)}(t,z)$ has a power series of the form

$$b^{(g)}(t,z) = \sum_{r=0}^{\infty}(2^r r!)^{g-1}L(t,z)^r \qquad \text{where} \quad L(t,z) = \frac{zt}{1-t(1-z)}\ . \tag{97}$$

This satisfies

$$2^{g-1}L(t,z)\left(1 + \frac{L(t,z)}{\partial L/\partial t}\frac{\partial}{\partial t}\right)^{g-1} b^{(g)}(t,z) + 1 = b^{(g)}(t,z)\ . \tag{98}$$

Plugging $b^{(g)}(t,z) = \sum_{n=0}^{\infty} b^{(g)}_n(z)\, t^n$ into this leads to a recursion for $b^{(g)}_n(z)$ by standard series methods. Using $\frac{\partial}{\partial z}$ instead of $\frac{\partial}{\partial t}$ leads to a mixed differential/recursion equation.

A similar method can be used to derive recurrence or mixed differential/recurrence equations in the circular case. Instead of using $a^{(g)}(t,z) = \sum_{n=0}^{\infty} a^{(g)}_n(z)\, t^n$ and $b^{(g)}(t,z) =$

$\sum_{n=0}^{\infty} b_n^{(g)}(z)\, t^n$ in the above procedure, use these formulas that follow from (82) and (79), to express the linear generating functions in terms of the circular generating functions:

$$a^{(g)}(t,z) = \frac{\sum_{n=0}^{\infty} \mathring{a}_n^{(g)}(z)\, t^n \quad + \quad \widetilde{G}zt(1+zt) - 1/(1-t(1-z))}{\widetilde{G}zt + zt/(1-t(1-z))^2)} \tag{99}$$

$$b^{(g)}(t,z) = \frac{\sum_{n=0}^{\infty} \mathring{b}_n^{(g)}(z)\, t^n \quad - \quad 1/(1-t(1-z))}{zt/(1-t(1-z))^2}\ . \tag{100}$$

A Technical proofs

A.1 Proof of Theorem 3.2: asymptotic number of signed arrangements

Proof of Theorem 3.2. We will use the falling factorial notation

$$(x)^{\underline{k}} = x(x-1)(x-2)\cdots(x-k+1)\ . \tag{101}$$

By (3),

$$\frac{b_{n,n-q}^{(g)}}{(2^{n-q}(n-q)!)^{g-1}\binom{n-1}{n-q-1}} = \sum_{r=1}^{n-q}(-1)^{n-q-r}\binom{n-q-1}{r-1}\left(\frac{2^r r!}{2^{n-q}(n-q)!}\right)^{g-1}$$

Change r to $n-q-r$:

$$= \sum_{r=0}^{n-q-1}(-1)^r\binom{n-q-1}{n-q-r-1}\left(\frac{2^{n-q-r}(n-q-r)!}{2^{n-q}(n-q)!}\right)^{g-1}$$

$$= \sum_{r=0}^{n-q-1}(-\tfrac{1}{2})^r\frac{(n-q-1)!}{(n-q-r-1)!\,r!}\left(\frac{(n-q-r)!}{(n-q)!}\right)^{g-1}$$

Pull off one factor of $(n-q-r)!/(n-q)!$:

$$= \sum_{r=0}^{n-q-1}\frac{(-\frac{1}{2})^r}{r!}\frac{n-q-r}{n-q}\left(\frac{(n-q-r)!}{(n-q)!}\right)^{g-2}$$

$$= \sum_{r=0}^{n-q-1}\frac{(-\frac{1}{2})^r}{r!}\left(1-\frac{r}{n-q}\right)\left(\frac{1}{(n-q)^{\underline{r}}}\right)^{g-2} \tag{102}$$

We treat the cases $g=2$ and $g>2$ separately. For $g>2$, we have

$$\frac{b_{n,n-q}^{(g)}}{(2^{n-q}(n-q)!)^{g-1}\binom{n-1}{n-q-1}} = \sum_{r=0}^{n-q-1}\frac{(-\frac{1}{2})^r}{r!}\left(1-\frac{r}{n-q}\right)\left(\frac{1}{(n-q)^{\underline{r}}}\right)^{g-2}$$

Term $r=0$ is 1. The terms have alternating signs and decreasing magnitudes (take ratios) that go to 0 for each $r>0$ as $n\to\infty$. So the sum goes to 1 as $n\to\infty$, giving the $g>2$ case of (4).

For $g = 2$, Eq. (102) simplifies to

$$\begin{aligned}
\frac{b^{(2)}_{n,n-q}}{2^{n-q}\,(n-q)!\,\binom{n-1}{n-q-1}} &= \sum_{r=0}^{n-q-1} \frac{(-\frac{1}{2})^r}{r!}\left(1 - \frac{r}{n-q}\right) \\
&= \sum_{r=0}^{n-q-1} \frac{(-\frac{1}{2})^r}{r!} + \frac{1}{2(n-q)} \sum_{r=1}^{n-q-1} \frac{(-\frac{1}{2})^{r-1}}{(r-1)!} \\
&= \exp_{n-q-1}(-\tfrac{1}{2}) + \frac{\exp_{n-q-2}(-\frac{1}{2})}{2(n-q)} \;\to\; \exp(-\tfrac{1}{2}) + 0 \quad \text{as } n \to \infty,
\end{aligned}$$

which gives the $g = 2$ case of (4). The ratio of the denominators in the limits in equations (4) and (5) is

$$\frac{2^{n-q}(n-q)!\,\binom{n-1}{n-q-1}}{2^n\, n!} = \frac{(n-q)!\,(n-1)!}{2^q n!\,(n-q-1)!\,q!} = \frac{n-q}{n} \cdot \frac{1}{2^q\, q!}$$

which goes to $1/(2^q\, q!)$ as $n \to \infty$, leading to (5). □

A.2 Proof of Theorem 5.9: asymptotic number of unsigned arrangements

The proof of Theorem 5.9 uses of a summation whose properties we list in the following lemma:

Lemma A.1. *Let $0 \le m \le n$ be integers, $G > 0$ be real, and define*

$$S(n, m, G) = \sum_{i=0}^{\min(m,n-m)} G^i \binom{n-m}{i} \binom{n-i-1}{n-m-1} \frac{m!(n-m)!}{n!} \,. \tag{103}$$

Note that all terms are nonnegative. The boundaries are $S(n, 0, G) = 1$ (for $n \ge 0$) and $S(n, n, G) = 0$ (for $n > 0$). Then

$$\begin{aligned}
S(n, m, G) &= (G + 1)^m (1 + O(m^2/n)) && \textit{as } n \to \infty; && (104)\\
S(n, m + 1, G)/S(n, m, G) &< G + 1 && \textit{for } 0 \le m < n. && (105)
\end{aligned}$$

The proof of the lemma is at the end of this section.

Proof of Theorem 5.9. Let $n > 0$, plug $k = n - q$ into (37), divide by $n!(n - q)!^{g-2}$, and judiciously insert a factor of $1 = (n - r)!/(n - r)!$:

$$\begin{aligned}
\frac{a^{(g)}_{n,n-q}}{n!(n-q)!^{g-2}} = \sum_{r=1}^{n-q} \Bigg(\left(\frac{r!}{(n-q)!}\right)^{g-2} \frac{(-1)^{n-q-r}}{(n-r)!} \binom{n-r}{n-q-r} \\
\cdot \sum_{i=0}^{\min(r,n-r)} G^i \binom{r}{i} \binom{n-i-1}{r-1} \frac{(n-r)!\,r!}{n!} \Bigg) \,.
\end{aligned} \tag{106}$$

Change the summation variable from r to $m = n - r$ and use function (103):

$$= \sum_{m=q}^{n-1} \left(\frac{(n-m)!}{(n-q)!} \right)^{g-2} \frac{(-1)^{m-q}}{m!} \binom{m}{q} S(n,m,G) \ . \tag{107}$$

This is an alternating sum. Using (105), the ratio of term $m+1$ over term m has absolute value bounded by

$$\frac{G+1}{(n-m)^{g-2}(m+1-q)} \ . \tag{108}$$

The proof splits into two cases, $g = 2$ and $g > 2$.

For $g > 2$, for all m the ratio (108) tends to 0 as $n \to \infty$. So as $n \to \infty$, (107) is dominated by its first term $S(n,q,G) = ((G+1)^q/q!)(1 - O(q^2/n)) = (2^{q(g-1)}/q!)(1 - O(q^2/n))$ with error bounded by its second term, which is a fraction $O(1/(n-q)^{g-2})$ of the first term. This rearranges to the $g > 2$ case of (39).

For $g = 2$ we have $G = 1$. Ratio (108) simplifies to $(G+1)/(m+1-q) = 2/(m+1-q)$, which is below 1 when $m \geq q+2$. So if we truncate the sum (107) at any term $m \geq q+2$, the error is bounded by the next term. We choose to terminate at $m = \lfloor n^{1/3} \rfloor$ so the error term is $O(m^2/n) = O(n^{-1/3})$. As $n \to \infty$, the truncated sum termwise approaches

$$\sum_{m=q}^{\lfloor n^{1/3} \rfloor} \frac{(-1)^{m-q}}{m!} \binom{m}{q} (G+1)^m (1 + O(n^{-1/3}))$$
$$\to \sum_{m=q}^{\infty} \frac{(-1)^{m-q}}{m!} \binom{m}{q} (G+1)^m = \sum_{m=q}^{\infty} \frac{(-1)^{m-q}}{m!} \binom{m}{q} 2^m = \frac{2^q \exp(-2)}{q!} \ .$$

This gives (40) and the $g = 2$ case of (39). □

Now we prove Lemma A.1.

Proof of (104). The sum $S(n,m,G)$ in (103) may be rewritten using factorials and falling factorial notation (101):

$$S(n,m,G) = \sum_{i=0}^{\min(m,n-m)} G^i \frac{(n-m)!\,(n-i-1)!\,m!\,(n-m)!}{i!\,(n-m-i)!\,(n-m-1)!\,(m-i)!\,n!} \tag{109}$$

$$= \sum_{i=0}^{\min(m,n-m)} G^i \binom{m}{i} \frac{(n-m) \cdot (n-i-1)^{\underline{m-1}}}{(n)^{\underline{m}}} \tag{110}$$

Note that $(n-m) \cdot (n-i-1)^{\underline{m-1}}/(n)^{\underline{m}}$ is a ratio of two monic polynomials in n, each of degree m. As $n \to \infty$, it asymptotically equals $1 - (i \cdot m + m - i)/n + O(1/n^2)$. Also, $\min(m, n-m) = m$ when $n \geq 2m$, and $= n - m$ when $n < 2m$. So as $n \to \infty$, (110) is

$$\sum_{i=0}^{m} G^i \binom{m}{i} \left(1 - \frac{i \cdot m + m - i}{n} + O(\tfrac{1}{n^2}) \right) = (G+1)^m \cdot \left(1 + O(\tfrac{m^2}{n})\right) \ . \quad \square$$

Proof of (105). If $0 = m < n$ then $S(n,1,G)/S(n,0,G) = (1-\frac{1}{n})(G+1)/1 < G+1$.

Let $0 < m < n$. Set $h = n - m > 0$ and $m' = \min(m, n-m)$. Consider

$$D = (G+1)S(n,m,G) - S(n,m+1,G) \tag{111}$$

$$= T + \sum_{i=0}^{m'} G^i \binom{m}{i} \frac{h \cdot (n-i-1)^{\underline{m-1}}}{(m+1-i)(n)^{\underline{m}}} \left(\frac{i(m-1)}{h+1-i} + \frac{(m+1)(h+i(h-1))}{h^2} \right) \tag{112}$$

where $T = 0$ if $n < 2m+2$, and if $n \geq 2m+2$, there is an additional term

$$T = G^{m+1} \frac{(n-m-1)^{\underline{m+1}}}{(n)^{\underline{m+1}}} \left(\frac{(m+1)^2}{n-2m-1} - \frac{m^2}{n-2m} \right) . \tag{113}$$

The coefficient of G^i in (112) is positive for $i = 0, \ldots, m'$:

- $h = n - m > 0$ since $m < n$. This also gives $(m+1)(h+i(h-1)) > 0$ and $h^2 > 0$.
- $i \leq n - m = h$ and $i \leq m$, so $(n-i-1)^{\underline{m-1}} > 0$; $h+1-i > 0$; and $m+1-i > 0$.
- $m \leq n-1$ so $(n)^{\underline{m}} > 0$.

If $n \geq 2m+2$, we must also show that $T > 0$. Since $(m+1)^2/(n-2m-1)$ has a larger numerator and smaller denominator than $m^2/(n-2m)$, the difference is positive and $T > 0$.

Thus we have shown $D > 0$, which proves (105).

Note that $S(n,0,G) = 1$, so iterating (105) gives $S(n,m,G) < (1+G)^m$ for $1 \leq m \leq n$. Thus, the $O(m^2/n)$ error term in (104) as $n \to \infty$ is an overestimate as m increases, but it is sufficient for our purposes. $\square$

A.3 Proof of Theorem 9.8: weight generating function for unsigned circular arrangements

This appendix section builds to a proof of Theorem 9.8. For this section, define

$$P = 1 + U(t) , \qquad Q = \widetilde{G}\, U_1\, t , \qquad R = 1 + V(t) , \qquad S = G\, V_1\, t . \tag{114}$$

Lemma A.2.

$$\phi^{-1}(R) = (1-PQ)^{-1}P = P(1-QP)^{-1} \tag{115}$$

$$\phi^{-1}(S) = -Q \tag{116}$$

$$\phi^{-1}(R^{-1}) = P^{-1} - Q = P^{-1}(1-PQ) = (1-QP)P^{-1} \tag{117}$$

$$\phi^{-1}\left((1-RS)^{-1}\right) = 1 - PQ \tag{118}$$

$$\phi^{-1}(V'(t)) = (1-PQ)^{-1}\left(U'(t) + \tfrac{PQP}{t}\right)(1-QP)^{-1} \tag{119}$$

Proof. Eqs. (115), (116), and (117) follow from (30), (29) and (31) respectively. For (118):

$$\phi^{-1}(1-RS) = 1 - \big((1-PQ)^{-1}P\big)(-Q) = 1 + (1-PQ)^{-1}PQ = (1-PQ)^{-1}\ .$$

For (119), rearrange (115) to $(1-PQ)\,\phi^{-1}(R) = P$. Differentiate with respect to t:

$$(-P'Q - PQ')\,\phi^{-1}(R) + (1-PQ)\phi^{-1}(R') = P'$$

Rearrange it and plug in $\phi^{-1}(R) = P(1-QP)^{-1}$:

$$(1-PQ)\phi^{-1}(R') = P' + (P'Q+PQ')\,\phi^{-1}(R)P' + (P'Q+PQ')\,P(1-QP)^{-1}$$

Plug in $P' = U'(t)$, $Q' = Q/t$, and $R' = V'(t)$, and simplify:

$$\begin{aligned}(1-PQ)\phi^{-1}(V'(t)) &= U'(t) + \big(U'(t)Q + PQ/t\big)\,P(1-QP)^{-1}\\ &= U'(t)\big(1 + QP(1-QP)^{-1}\big) + \tfrac{PQP}{t}(1-QP)^{-1}\\ &= U'(t)(1-QP)^{-1} + \tfrac{PQP}{t}(1-QP)^{-1}\\ &= \big(U'(t) + \tfrac{PQP}{t}\big)(1-QP)^{-1}\end{aligned}$$

Left multiply by $(1-PQ)^{-1}$ to get (119). □

Next we make a generating function version of Theorem 9.7:

Lemma A.3. *(i) Set* $W(t) = \sum_{n=1}^{\infty} t^n W_n$, *so* $\phi(\mathring{U}(t)) = \mathring{V}(t) + W(t)$. *Then*

$$W(t) = \Big[\big(1+V(t)+t\,V'(t)\big)GV_1t\big(1-\big(1+V(t)\big)GV_1t\big)^{-1} - GV_1t - G(G+1)V_1V_1t^2\Big]_{\circ} \tag{120}$$

(ii) $\phi^{-1}(\mathring{V}(t)) = \mathring{U}(t) - [\phi^{-1}(W(t))]_{\circ}$, *where*

$$-[\phi^{-1}(W(t))]_{\circ} = \Big[\big(1-(1+U(t))\widetilde{G}U_1t\big)^{-1}\big(1+U(t)+tU'(t)\big)\widetilde{G}U_1t - \widetilde{G}U_1t - \widetilde{G}U_1U_1t^2\Big]_{\circ}$$

Note: Compare the equations for $\phi(\mathring{U}(t)) = \mathring{V}(t) + W(t)$ and $\phi^{-1}(\mathring{V}(t)) = \mathring{U}(t) + [-\phi^{-1}(W(t))]_{\circ}$ in Lemma A.3. They are almost dual in the sense of Theorem 5.5(vi), except for (i) the coefficient of t^2; (ii) we must straighten the monomials; (iii) in the case of single chromosomes treated in this manuscript, $\mathring{U}_n$ and $\mathring{V}_n$ only appear in functions linearly; they are never multiplied by non-constant expressions. We may extend Theorem 5.5(vi) by adding duals for $n \neq 2$: $\phi(\mathring{U}_n) \leftrightarrow \phi^{-1}(\mathring{V}_n)$ and $\mathring{V}_n \leftrightarrow \mathring{U}_n$. For $n=2$, $\phi(\mathring{U}_2) - (\mathring{V}_2 + GV_1V_1) \leftrightarrow \phi^{-1}(\mathring{V}_2) - (\mathring{U}_2 + \widetilde{G}(\widetilde{G}+1)U_1U_1)$.

Proof of Lemma A.3. (i) Eq. (71) does not apply to $n = 1, 2$; to make it apply to all $n \geq 0$, rewrite the left side of that equation as $W_n + E_n$, where E_n is a correction term: $E_0 = 0$, $E_1 = GV_1$, $E_2 = G(G+1)V_1V_1$, and $E_n = 0$ for $n > 3$. Then for all $n \geq 0$,

$$W_n + E_n = \sum_{k=1}^{n} \sum_{\beta\in\mathcal{C}_{n,k}} G^k \beta_1 \left[V_{\beta_1-1}V_1V_{\beta_2-1}V_1\cdots V_{\beta_k-1}V_1\right]_{\circ} \tag{121}$$

Multiply by t^n and sum over $n \geq 0$. Then

$$W(t) + E(t) = \left[\sum_{k=1}^{\infty} \left(\sum_{\beta_1=1}^{\infty} G\beta_1 V_{\beta_1-1} V_1 t^{\beta_1} \right) \left(\sum_{\beta_2=1}^{\infty} G V_{\beta_2-1} V_1 t^{\beta_2} \right)^{k-1} \right]_{\circ} \tag{122}$$

where $E(t) = \sum_{n=0}^{\infty} E_n t^n = GV_1 t + G(G+1)V_1V_1t^2$. The β_1 sum in (122) is

$$\begin{aligned}
\sum_{\beta_1=1}^{\infty} G\beta_1 V_{\beta_1-1} V_1 t^{\beta_1} &= \left(\sum_{\beta_1=1}^{\infty} \beta_1 V_{\beta_1-1} t^{\beta_1-1} \right) GV_1 t \\
&= \Big(t\big(1+V(t)\big)\Big)' GV_1 t = \big(1 + V(t) + t\,V'(t)\big) GV_1 t
\end{aligned}$$

and the β_2 sum in (122) is $\big(1+V(t)\big)V_1 Gt$. Plugging these into (122) and evaluating the geometric series (sum over k) gives

$$W(t) + E(t) = \Big[\big(1 + V(t) + t\,V'(t)\big)GV_1 t\big(1 - \big(1+V(t)\big)V_1 Gt\big)^{-1}\Big]_{\circ} \tag{123}$$

(ii) Using the abbreviations (114), we rewrite (120) as

$$W(t) = [(R + t\,V'(t))S(1-RS)^{-1} - S + S^2/\widetilde{G}]_{\circ} \tag{124}$$

where we used $-(G+1)/G = 1/\widetilde{G}$. Now compute ϕ^{-1} of this equation piece-by-piece:

$$\begin{aligned}
\phi^{-1}\big(R + t\,V'(t)\big) &= P(1-QP)^{-1} + t(1-PQ)^{-1}\big(U'(t) + \tfrac{PQP}{t}\big)(1-QP)^{-1} \\
&= \big(P + (1-PQ)^{-1}(t\,U'(t) + PQP)\big)(1-QP)^{-1}\ .
\end{aligned}$$

Right-multiply by $\phi(S(1-RS)^{-1}) = (-Q)(1-PQ) = -Q + QPQ = -(1-QP)Q$:

$$\begin{aligned}
&\phi^{-1}\big((R + t\,V'(t))S(1-RS)^{-1}\big) \\
&\qquad = -\big(P + (1-PQ)^{-1}(t\,U'(t) + PQP)\big)(1-QP)^{-1}(1-QP)Q \\
&\qquad = -\big(P + (1-PQ)^{-1}(t\,U'(t) + PQP)\big)Q \\
&\qquad = -(1-PQ)^{-1}t\,U'(t)Q - PQ - (1-PQ)^{-1}PQPQ \\
&\qquad = -(1-PQ)^{-1}t\,U'(t)Q - (1-PQ)^{-1}PQ \\
&\qquad = -(1-PQ)^{-1}(t\,U'(t) + P)Q\ .
\end{aligned}$$

The remaining terms in (124) are $\phi^{-1}(-S + S^2/\widetilde{G}) = Q + Q^2/\widetilde{G}$. Then

$$\begin{aligned}
\phi^{-1}(W(t)) &= [-(1-PQ)^{-1}(t\,U'(t) + P)Q + Q + Q^2/\widetilde{G}]_{\circ} && (125) \\
&= [-(1-(1+U(t))\widetilde{G}U_1t)^{-1}(1+U(t)+tU'(t))\widetilde{G}U_1t && (126) \\
&\qquad + \widetilde{G}U_1t + \widetilde{G}U_1U_1t^2]_{\circ} && \square
\end{aligned}$$

Proof of Theorem 9.8. By Theorem 9.7, we have $\mathring{A}^{(g)}(\vec{U};\mathring{\vec{U}};t) = [\phi^{-1}\big(\mathring{B}^{(g)}(\vec{V};\mathring{\vec{V}};t)\big)]_\circ$, where we use (114) and then (124) to rewrite (67) as

$$\begin{aligned}\mathring{B}^{(g)}(\vec{V};\mathring{\vec{V}};t) &= \Big[(\mathring{V}(t)+W) - W + t\,V'(t)\left(-R^{-1}+R^{-1}B^{(g)}(\vec{V};t)R^{-1}\right)\Big]_\circ \\ &= \Big[(\mathring{V}(t)+W) + t\,V'(t)\left(-S(1-RS)^{-1} - R^{-1}+R^{-1}B^{(g)}(\vec{V};t)R^{-1}\right) \\ &\qquad - RS(1-RS)^{-1} + S - S^2/\widetilde{G}\Big]_\circ \end{aligned} \tag{127}$$

Apply ϕ^{-1} to this piece-by-piece, and straighten. The left side is $[\phi^{-1}(\mathring{B}^{(g)}(\vec{V};\mathring{\vec{V}};t))]_\circ = \mathring{A}^{(g)}(\vec{U};\mathring{\vec{U}};t)$. The right side splits into three parts, $T_1+T_2+T_3$. The first part is $T_1 = [\phi^{-1}(\mathring{V}(t)+W(t))]_\circ = \mathring{U}(t)$. The last part is

$$\begin{aligned} T_3 &= \phi^{-1}\big(-RS(1-RS)^{-1} + S - S^2/\widetilde{G}\big) \\ &= (1-PQ)^{-1}PQ(1-PQ) - Q - Q^2/\widetilde{G} \\ &= PQ - Q - Q^2/\widetilde{G} = (P - Q/\widetilde{G} - 1)Q = (U(t)-U_1t)\widetilde{G}U_1t \end{aligned}$$

The middle part, T_2, involves $t\,V'(t)$ times a large expression. Note that $-S(1-RS)^{-1}) - R^{-1} = -R^{-1}(RS(1-RS)^{-1}+1) = -R^{-1}(1-RS)^{-1}$. Thus

$$\begin{aligned} T_2 &= \phi^{-1}\big(t\,V'(t)\big(-R^{-1}(1-RS)^{-1} + R^{-1}B^{(g)}(\vec{V};t)R^{-1}\big)\big) \\ &= \phi^{-1}(t\,V'(t)R^{-1})\phi^{-1}\big(-(1-RS)^{-1} + B^{(g)}(\vec{V};t)R^{-1}\big)\,. \end{aligned}$$

The first factor of T_2 evaluates to

$$\begin{aligned} \phi^{-1}(t\,V'(t)R^{-1}) &= (1-PQ)^{-1}(t\,U'(t)+PQP)(1-QP)^{-1}(1-QP)P^{-1} \\ &= (1-PQ)^{-1}(t\,U'(t)+PQP)P^{-1}\,. \end{aligned}$$

The second factor of T_2 evaluates to

$$-(1-PQ) + A^{(g)}(\vec{U};t)P^{-1}(1-PQ) = \big(-1+A^{(g)}(\vec{U};t)P^{-1}\big)(1-PQ)\,.$$

So $T_2 = (1-PQ)^{-1}(t\,U'(t)+PQP)P^{-1}\big(-1+A^{(g)}(\vec{U};t)P^{-1}\big)(1-PQ)$. When we straighten this, the leading $(1-PQ)^{-1}$ and trailing $(1-PQ)$ cancel, and we further simplify:

$$\begin{aligned} [T_2]_\circ &= [(t\,U'(t)+PQP)P^{-1}\big(-1+A^{(g)}(\vec{U};t)P^{-1}\big)]_\circ \\ &= [(t\,U'(t)+PQP)P^{-1}(A^{(g)}(\vec{U};t)-P)P^{-1}]_\circ \\ &= [P^{-1}(t\,U'(t)+PQP)P^{-1}(A^{(g)}(\vec{U};t)-P)]_\circ \\ &= [(t\,P^{-1}U'(t)P^{-1}+Q)(A^{(g)}(\vec{U};t)-P)]_\circ \end{aligned}$$

Combining all three pieces gives

$$\begin{aligned} \mathring{A}^{(g)}(\vec{U};\mathring{\vec{U}};t) &= [T_1+T_2+T_3]_\circ \\ &= \mathring{U}(t) + \Big[(t\,P^{-1}U'(t)P^{-1}+Q)(A^{(g)}(\vec{U};t)-P) + (U(t)-U_1t)\widetilde{G}U_1t\Big]_\circ \end{aligned}$$

Replacing P,Q by their definitions (114) gives (73). $\square$

References

[1] Morton Abramson and W. O. J. Moser, *Permutations without rising or falling w-sequences*, Ann. Math. Statist. **38** (1967), 1245–1254.

[2] George E. Andrews, *The theory of partitions*, Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1976, Encyclopedia of Mathematics and its Applications, Vol. 2.

[3] George E. Andrews and Kimmo Eriksson, *Integer partitions*, Cambridge University Press, Cambridge, 2004.

[4] D.A. Bader, B.M.E. Moret, and M. Yan, *A linear-time algorithm for computing inversion distances between signed permutations with an experimental study*, J. Comput. Biol **8** (2001), no. 5, 483–491.

[5] V. Bafna and P.A. Pevzner, *Genome rearrangements and sorting by reversals*, Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science, 1993, pp. 148–157.

[6] ______, *Genome rearrangements and sorting by reversals*, SIAM Journal on Computing **25** (1996), 272–289.

[7] Nathanaël Berestycki and Rick Durrett, *A phase transition in the random transposition random walk*, Probab. Theory Related Fields **136** (2006), no. 2, 203–233.

[8] G. Bourque and P.A. Pevzner, *Genome-scale evolution: reconstructing gene orders in the ancestral species*, Genome Research **12** (2002), 9748–9753.

[9] D.R. Cox, M. Burmeister, E.R. Price, S. Kim, and R.M. Myers, *Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes*, Science **250** (1990), no. 4978, 245–250.

[10] Leonard Eugene Dickson, *History of the theory of numbers. Vol. II: Diophantine analysis*, Cargnegie Institution of Washington, Washington, D.C., 1920.

[11] T. Dobzhansky and A. Sturtevant, *Inversions in the chromosomes of* Drosophila pseudoobscura, Genetics **23** (1938), 28–64.

[12] S. Hannenhalli and P.A. Pevzner, *Transforming men into mice (polynomial algorithm for genomic distance problem)*, 36th Annual Symposium on Foundations of Computer Science (Milwaukee, WI, 1995), IEEE Comput. Soc. Press, Los Alamitos, CA, 1995, pp. 581–592.

[13] ______, *To cut ... or not to cut (applications of comparative physical maps in molecular evolution)*, Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996) (New York), ACM, 1996, pp. 304–313.

[14] ______, *Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals*, J. ACM **46** (1999), no. 1, 1–27.

[15] Irving Kaplansky, *The asymptotic distribution of runs of consecutive elements*, Ann. Math. Statistics **16** (1945), 200–203.

[16] J. Kececioglu and D. Sankoff, *Exact and approximation algorithms for the inversion distance between two chromosomes*, Combinatorial pattern matching (Padova, 1993), Springer, Berlin, 1993, pp. 87–105.

[17] ———, *Efficient bounds for oriented chromosome inversion distance*, Combinatorial pattern matching (Asilomar, CA, 1994), Springer, Berlin, 1994, pp. 307–325.

[18] ———, *Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement*, Algorithmica **13** (1995), no. 1-2, 180–210.

[19] William J. Murphy, Denis M. Larkin, Annelie Everts van der Wind, Guillaume Bourque, Glenn Tesler, and et al., *Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps*, Science **309** (2005), 613–617.

[20] J.H. Nadeau and B.A. Taylor, *Lengths of chromosomal segments conserved since divergence of man and mouse*, Proceedings of the National Academy of Sciences USA **81** (1984), 814–818.

[21] J.D. Palmer and L.A. Herbon, *Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence*, Journal of Molecular Evolution **27** (1988), 87–97.

[22] J.D. Palmer, B. Osorio, and W.F. Thompson, *Evolutionary significance of inversions in legume chloroplast DNAs*, Current Genetics **14** (1988), 65–74.

[23] Qian Peng, Pavel A Pevzner, and Glenn Tesler, *The fragile breakage versus random breakage models of chromosome evolution*, PLoS Comput Biol **2** (2006), no. 2, e14.

[24] P. Pevzner and G. Tesler, *Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution*, Proceedings of the National Academy of Sciences USA **100** (2003), no. 13, 7672–7677.

[25] P.A. Pevzner and G. Tesler, *Genome rearrangements in mammalian evolution: Lessons from human and mouse genomes*, Genome Research **13** (2003), no. 1, 37–45.

[26] P.A. Pevzner and Glenn Tesler, *Transforming men into mice: the Nadeau-Taylor chromosomal breakage model revisited*, Proceedings of the Seventh Annual International Conference on Research in Computational Molecular Biology (RECOMB), 2003.

[27] John Riordan, *Inverse relations and combinatorial identities*, Amer. Math. Monthly **71** (1964), 485–498.

[28] ———, *A recurrence for permutations without rising or falling successions*, Ann. Math. Statist. **36** (1965), 708–710.

[29] David Sankoff and Phil Trinh, *Chromosomal breakpoint re-use in the inference of genome sequence rearrangement*, Proceedings of the eighth annual international conference on computational molecular biology (2004), 30–35.

[30] ———, *Chromosomal breakpoint re-use in genome sequence rearrangement*, Journal of Computational Biology **12** (2005), no. 6, 812–821.

[31] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, 2008, http://www.research.att.com/ñjas/sequences/A002464.

[32] G.A. Watterson, W.J. Ewens, T.E. Hall, and A. Morgan, *The chromosome inversion problem*, Journal of Theoretical Biology **99** (1982), 1–7.

[33] J. Wolfowitz, *Additive partition functions and a class of statistical hypotheses*, Ann. Math. Statistics **13** (1942), 247–279.

[34] ______, *Note on runs of consecutive elements*, Ann. Math. Statistics **15** (1944), 97–98.

[35] Wei Xu, Chunfang Zheng, and David Sankoff, *Paths and cycles in breakpoint graphs of random multichromosomal genomes*, RECOMB 2006 International Workshop, RECOMB-CG 2006 Montreal, Canada, September 24-26, 2006 Proceedings (Berlin, Germany), Lecture Notes in Bioinformatics, vol. 4205, Springer Verlag, 2006, pp. 51–62.