Multiresolution Analysis of Incomplete Rankings

Stéphan Clémençon, Jérémie Jakubowicz, Eric Sibony

Multiresolution Analysis of Incomplete Rankings *

Stéphan Clémençon[1], Jérémie Jakubowicz[2], and Eric Sibony[†1]

[1]LTCI UMR No. 5141 Telecom ParisTech/CNRS, Institut Mines-Telecom, Paris, 75013, France
[2]SAMOVAR UMR No. 5157 Telecom SudParis/CNRS, Institut Mines-Telecom, Paris, 75013, France

Abstract

Incomplete rankings on a set of items $\{1, \ldots, n\}$ are orderings of the form $a_1 \prec \cdots \prec a_k$, with $\{a_1, \ldots a_k\} \subset \{1, \ldots, n\}$ and $k < n$. Though they arise in many modern applications, only a few methods have been introduced to manipulate them, most of them consisting in representing any incomplete ranking by the set of all its possible linear extensions on $\{1, \ldots, n\}$. It is the major purpose of this paper to introduce a completely novel approach, which allows to treat incomplete rankings directly, representing them as injective words over $\{1, \ldots, n\}$. Unexpectedly, operations on incomplete rankings have very simple equivalents in this setting and the topological structure of the complex of injective words can be interpretated in a simple fashion from the perspective of ranking. We exploit this connection here and use recent results from algebraic topology to construct a multiresolution analysis and develop a wavelet framework for incomplete rankings. Though purely combinatorial, this construction relies on the same ideas underlying multiresolution analysis on a Euclidean space, and permits to localize the information related to rankings on each subset of items. It can be viewed as a crucial step toward nonlinear approximation of distributions of incomplete rankings and paves the way for many statistical applications, including preference data analysis and the design of recommender systems.

Keywords. Incomplete rankings, Multiresolution Analysis, Wavelets, Injective Words.

1 Introduction

Data expressing rankings or preferences have become ubiquitous in the Big Data era. Operating continuously on still more content, modern applications such as recommendation systems and search engines generate and/or exploit massive data of this nature. The design of statistical machine-learning algorithms, tailored to this type of data, is crucial to optimize the performance of such systems (*e.g.* rank documents by degree of relevance for a specific query in information retrieval, propose a sorted list of items/products to a prospect she/he is most liable to buy in e-commerce).

*This work was supported by Agence Nationale de la Recherche (France) grant ANR-11-IDEX-0003-02.

†Corresponding author - email: eric.sibony@telecom-paristech.fr - postal address: Telecom ParisTech 37-39 rue Dareau, 75014 Paris, France.

A well studied situation is when raw data are of the form of "full rankings" on a given set of items indexed by $[\![n]\!] = \{1, \ldots, n\}$ and are then described by permutations σ on $[\![n]\!]$ that map an item to its rank, $a_1 \prec \ldots \prec a_n$ with $a_i = \sigma^{-1}(i)$ for $i \in [\![n]\!]$. The variability of observations is represented by a probability distribution p on the set $\mathfrak{S}_n$ of all the permutations on $[\![n]\!]$, which can be seen as an element of the space

$$L(\mathfrak{S}_n) = \{f : \mathfrak{S}_n \to \mathbb{R}\},$$

such that $p(\sigma) \geq 0$ for all $\sigma \in \mathfrak{S}_n$ and $\sum_{\sigma \in \mathfrak{S}_n} p(\sigma) = 1$. Though empirical estimation of p may appear as a problem of disarming simplicity at first glance, it is actually a great statistical challenge because the number of possible rankings (*i.e.* $\mathfrak{S}_n$'s cardinality) explodes as $n!$ with the number of instances to be ranked. Traditional methods in machine-learning and statistics quickly become either intractable or inaccurate in practice and many approaches have been proposed these last few years to deal with preference data and overcome these challenges in different situations (*e.g.* [9], [16], [28], [13], [24], [14], [39]). Whatever the type of task considered (supervised, unsupervised), machine-learning algorithms generally rest upon the computation of statistical quantities such as averages or medians, summarizing/representing efficiently the data or the performance of a decision rule candidate applied to the data. However, summarizing ranking variability is far from straightforward and extending simple concepts such as an average or a median in the context of preference data raises a certain number of deep mathematical and computational problems, see [2], and call for new constructions.

One approach, much documented in the literature, consists in exploiting the algebraic structure of the (noncommutative) group $\mathfrak{S}_n$ and perform a harmonic analysis on $L(\mathfrak{S}_n)$, see for example [6], [37], [22], [19], [21]. This corresponds to a decomposition of the form

$$L(\mathfrak{S}_n) \cong \bigoplus_{\lambda} d_\lambda S^\lambda,$$

where the S^λ's are irreducible spaces invariant under the action of the translations $f \mapsto f(\sigma_0^{-1}.)$ for all $\sigma_0 \in \mathfrak{S}_n$, the λ's correspond to "frequencies", and the d_λ's are positive integers. The sign $\cong$ above means that the two spaces are isomorphic, the spaces S^λ being not necessarily subspaces of $L(\mathfrak{S}_n)$. This decomposition allows to localize the different spectral components of any function $f \in L(\mathfrak{S}_n)$. Furthermore, it is possible to define a (partial) order on the λ's that indicates the different level of "smoothness" of the elements of the corresponding S^λ's (for instance, the smoothest component is the space of constant functions), thus providing a natural framework for linear approximation in $L(\mathfrak{S}_n)$, see [15] or [18]. This framework also extends to the analysis of full rankings with ties, referred to as *bucket orders* (or *partial rankings* sometimes): for $1 \leq r \leq n$ and $\mu = (\mu_1, \ldots, \mu_r) \in \mathbb{N}^{*r}$ such that $\mu_1 + \cdots + \mu_r = n$, orderings of the type $a_{1,1}, \ldots, a_{1,\mu_1} \prec \ldots \prec a_{r,1}, \ldots, a_{r,\mu_r}$ described by mappings $\sigma : [\![n]\!] \to \{1, \ldots, r\}$ such that $\sigma^{-1}(\{i\}) = \{a_{i,1}, \ldots, a_{i,\mu_i}\}$ for any $i \in \{1, \ldots, r\}$. Among bucket orders, top-k rankings received special attention. They correspond to orderings of the form $a_1 \prec \cdots \prec a_k \prec$ *the rest*. The same notion of translation can be defined on the space M^μ of real-valued functions on partial rankings with fixed form $\mu = (\mu_1, \ldots, \mu_r)$, leading to a similar decomposition, called Young's rule,

$$M^\mu \cong \bigoplus_{\lambda} K_{\lambda,\mu} S^\lambda,$$

where the S^λ's are the same as before and the $K_{\lambda,\mu}$'s are integers ≥ 0 called the Kotska numbers.

This "$\mathfrak{S}_n$-based" harmonic analysis is however not suited for the analysis of ranked data of the form $a_1 \prec \cdots \prec a_k$ with $k < n$, *i.e.* when the rankings do not involve all the items. Such data shall be referred to as *incomplete rankings* throughout the article. Indeed, though [22] provides a remarkable application of $\mathfrak{S}_n$-based harmonic analysis to incomplete rankings, the decomposition into $\mathfrak{S}_n$-based translation-invariant components is by essence inadequate to localize the information relative to incomplete rankings on specific subsets of items. Yet incomplete rankings arise in many modern applications (such as recommending systems), where the number of objects to be ranked is very high whereas preferences are generally observed for a small number of objects only. In statistical signal and image processing, novel harmonic analysis tools such as wavelet bases and their extensions have recently revitalized structured data analysis and lead to sparse representations and efficient algorithms for a wide variety of statistical tasks: estimation, prediction, denoising, compression, clustering, etc. Inspired by advances in computational harmonic analysis and its applications to high-dimensional data analysis, our goal is to develop new concepts and algorithms to handle preference data taking the form of incomplete rankings, in order to solve statistical learning problems, motivated by the applications aforementioned, such as efficient/sparse representation of rankings, ranking aggregation, prediction of rankings. More precisely, it is the purpose of this paper to extend the principles of wavelet theory and construct a multiresolution analysis tailored for the description of incomplete rankings.

Let us introduce some preliminary notations to be more specific. For a finite set E of cardinality $|E|$ and $k \in \{0, \ldots, |E|\}$, we denote by $\binom{E}{k}$ the set of all subsets of E with k elements and we set $\mathcal{P}(E) = \bigcup_{j=2}^{|E|} \binom{E}{j}$. By definition, a ranking over a subset $A \in \mathcal{P}([\![n]\!])$ is described by a bijective mapping $\pi : A \to \{1, \ldots, |A|\}$ that assigns to each item $a \in A$ its rank (with respect to A). The ensemble $\mathfrak{S}'_A$ of such mappings can thus be viewed as the set of the incomplete rankings on $[\![n]\!]$ involving the items of A solely. Notice that unless $A = \{1, \ldots, k\}$ with $k \in \{2, \ldots, n\}$, this set is different from – yet in one-to-one correspondence with – the set $\mathfrak{S}_A$ of permutations on A, *i.e.* bijective mappings $\tau : A \to A$. The variability of incomplete rankings is then represented by a family $(P_A)_{A \in \mathcal{P}([\![n]\!])}$, where P_A is a probability distribution on $\mathfrak{S}'_A$. In order to guarantee that this family describes the distribution of the preferences of a statistical population, it is unavoidable to assume that the following "projectivity" property holds: for any $A = \{a_1, \ldots, a_k\} \in \mathcal{P}([\![n]\!])$ with $k < n$ and $b \in [\![n]\!] \setminus A$,

$$P_A(a_{i_1} \prec \ldots \prec a_{i_k}) = P_{A\cup\{b\}}(a_{i_1} \prec \ldots \prec a_{i_k} \prec b) + P_{A\cup\{b\}}(a_{i_1} \prec \ldots \prec b \prec a_{i_k}) + \\ \ldots + P_{A\cup\{b\}}(a_{i_1} \prec b \prec \ldots \prec a_{i_k}) + P_{A\cup\{b\}}(b \prec a_{i_1} \prec \ldots \prec a_{i_k}). \quad (*)$$

It simply means that the probability of a ranking should be conserved when a new item is added. It is straightforward to see that this assumption is equivalent to that stipulating the existence of a probability distribution p on $\mathfrak{S}_n$ such that for all $A \in \mathcal{P}([\![n]\!])$,

$$P_A(\pi) = \sum_{\sigma \in \mathfrak{S}_n(\pi)} p(\sigma),$$

where $\mathfrak{S}_n(\pi)$ is the set of all the permutations $\sigma \in \mathfrak{S}_n$ that extend π, *i.e.* such that for all $(a, b) \in A^2$, $\pi(a) < \pi(b) \Rightarrow \sigma(a) < \sigma(b)$. For a function $f \in L(\mathfrak{S}_n)$, we define its "marginal" on the subset $A \in \mathcal{P}([\![n]\!])$ by $f_A(\pi) = \sum_{\sigma \in \mathfrak{S}_n(\pi)} f(\sigma)$. Assumption $(*)$ then states that the P_A's are the marginals of a global probability distribution p on $\mathfrak{S}_n$. Now, in practical applications, incomplete rankings are not observed on all the subsets of $\mathcal{P}([\![n]\!])$ but only on a collection $\mathcal{A} \subset \mathcal{P}([\![n]\!])$, called

the observation design, and the variability of the observed incomplete rankings is represented by the sub-family $(P_A)_{A\in\mathcal{A}}$ of $(P_A)_{A\in\mathcal{P}([\![n]\!])}$. Defining the linear operators

$$\begin{aligned} M_A : L(\mathfrak{S}_n) &\to L(\mathfrak{S}'_A) \\ f &\mapsto f_A \end{aligned} \qquad \text{and} \qquad \begin{aligned} M_{\mathcal{A}} = \bigoplus_{A\in\mathcal{A}} M_A : L(\mathfrak{S}_n) &\to \bigoplus_{A\in\mathcal{A}} L(\mathfrak{S}'_A) \\ f &\mapsto (f_A)_{A\in\mathcal{A}}, \end{aligned}$$

the analysis of preference data must then be performed in the space

$$\mathbb{M}_{\mathcal{A}} = M_{\mathcal{A}}(L(\mathfrak{S}_n)).$$

Whereas the space $L(\mathfrak{S}_n)$ has been thoroughly studied, $\mathbb{M}_{\mathcal{A}}$ has never been investigated in contrast. Defining an explicit basis for this space or even simply calculating its dimension is indeed far from obvious. Furthermore, unless $\mathcal{A}$ is of the form $\bigcup_{j\in J}\binom{[\![n]\!]}{j}$ with $J \subset \{2,\dots,n\}$, $\mathfrak{S}_n$-based translations cannot be defined and $\mathfrak{S}_n$-based harmonic analysis cannot previously cannot be applied. Instead, one needs a decomposition that localizes the information related to each subset of items (which is by nature not invariant under $\mathfrak{S}_n$-based translations).

1.1 Main contributions

In this article, we construct for any $A \in \mathcal{P}([\![n]\!])$, the subspace W_A of $L(\mathfrak{S}_n)$ that localizes the information that is specific to marginals on A and not to marginals on other subsets. Denoting by ψ_0 the constant function in $L(\mathfrak{S}_n)$ equal to 1 and by $V^0 = \mathbb{R}\psi_0$ the subspace of constant functions, the major contribution of the present paper is to establish the linear decomposition

$$L(\mathfrak{S}_n) = V^0 \oplus \bigoplus_{B\in\mathcal{P}([\![n]\!])} W_B. \tag{1}$$

Notice that this decomposition is an equality and not an isomorphism, because the W_B's are subspaces of $L(\mathfrak{S}_n)$. Denoting by $\ker M$ the null space of any linear operator M, our construction of the spaces W_B then allows to localize the information of the marginal on any subset $A \in \mathcal{P}([\![n]\!])$ via

$$L(\mathfrak{S}_n) = \ker M_A \oplus \left[V^0 \oplus \bigoplus_{B\in\mathcal{P}(A)} W_B \right],$$

and more generally the information of the marginals on the subsets of any collection $\mathcal{A}$ via

$$L(\mathfrak{S}_n) = \ker M_{\mathcal{A}} \oplus \left[V^0 \oplus \bigoplus_{B\in\bigcup_{A\in\mathcal{A}}\mathcal{P}(A)} W_B \right].$$

This last decomposition gives the multiresolution decomposition of the space $\mathbb{M}_{\mathcal{A}}$

$$\mathbb{M}_{\mathcal{A}} = M_{\mathcal{A}}\left(V^0\right) \oplus \bigoplus_{B\in\bigcup_{A\in\mathcal{A}}\mathcal{P}(A)} M_{\mathcal{A}}\left(W_B\right), \tag{2}$$

where $M_{\mathcal{A}}\left(V^0\right)$ is the component related to constant functions and for each $B \in \bigcup_{A\in\mathcal{A}}\mathcal{P}(A)$, $M_{\mathcal{A}}\left(W_B\right)$ is the component that localizes the information specific to the marginal on B. Our result

relies on recent advances in algebraic topology about the homological structure of the complex of injective words established in [35]. We call the decomposition (1) a "multiresolution decomposition" because the subspaces localize meaningful parts of the global information of incomplete rankings at different "scales". We nonetheless draw attention on the fact that this decomposition is not orthogonal (as we shall see in section 4) and it is not a "multiresolution analysis" in the strict sense. Indeed, the discrete nature of $\mathfrak{S}_n$ does not allow to define any dilation operator. However, as shall be seen later in the paper, translation and "dezooming" operators can still be defined to reinforce the analogy between our construction and standard multiresolution analysis, see subsection 3.4.

In order to use this decomposition to perform approximation in $\mathbb{M}_{\mathcal{A}}$ in practice, one needs an explicit basis for each space $M_{\mathcal{A}}(W_B)$. The effective construction of such an explicit basis is far from being obvious, because each space W_B is defined by many linear constraints based on the complex combinatorial structure of $\mathfrak{S}_n$. However, this problem can be related to that of constructing a basis for the homology of certain types of simplicial complexes (namely boolean complexes of Coxeter systems), for which a solution was recently established in [33]. Here we adapt the results from [33] to exhibit an explicit basis Ψ_B for each space W_B. The concatenated family $\Psi = \{\psi_0\} \cup \bigcup_{B \in \mathcal{P}([\![n]\!])} \Psi_B$ is then a basis of $L(\mathfrak{S}_n)$ adapted to the multiresolution decomposition (1), which shall be referred to as a *wavelet basis* here. From the basis Ψ, one obtains, for any collection $\mathcal{A}$ of subsets of $[\![n]\!]$, the wavelet basis

$$\{M_{\mathcal{A}}(\psi_0)\} \cup \bigcup_{B \in \bigcup_{A \in \mathcal{A}} \mathcal{P}(A)} \{M_{\mathcal{A}}(\psi)\}_{\psi \in \Psi_B},$$

adapted to the multiresolution decomposition (2) of the space $\mathbb{M}_{\mathcal{A}}$. Again we draw attention on the fact that Ψ is not a wavelet basis in the strict sense, obtained from the dilations and translations of a "mother wavelet", because of the nature of decomposition (1). It happens however that the choice of the algorithm adapted from [33] to generate each Ψ_B for $B \in \mathcal{P}([\![n]\!])$ leads to a global structure for Ψ encoded in two general relations, strengthening the analogy with classic wavelet bases, see subsection 4.4.

1.2 Related work

To the best of our knowledge, only three approaches are documented in the literature to analyze incomplete rankings. The first method is based on the Luce-Plackett model (see [25], [32]), the sole parametric statistical model on the group of permutations that can be straightforwardly extended to incomplete rankings. It relies on a strong assumption, referred to as *Luce's choice axiom*, which reduces the complexity of the model, encapsulated by n parameters only (contrasting with the cardinality of $\mathfrak{S}_n$). It has been used in a wide variety of applications and several algorithms have been proposed to infer its parameters, see [17] or [1] for instance. Several numerical experiments on real datasets have shown however that its capacity to fit real data is limited, the model being too rigid to handle singularities observed in practice, refer to [29] and [39]. The two other approaches are non-parametric kernel methods. The one proposed in [22] is a diffusion kernel in the Fourier domain, and the one proposed in [39] is a triangular kernel with respect to the Kendall's tau distance. Though leading to efficient algorithms, both approaches deal with sets $\mathfrak{S}_n(\pi)$'s and not directly with incomplete rankings π's. This tends to blend the estimated probabilities of the incomplete rankings and thus induces a statistical bias. In contrast, our framework relies on the natural multiresolution structure of incomplete rankings and is the first to allow the definition of approximation procedures directly on this type of ranked data.

We point out that an alternative construction of a multiresolution analysis on $L(\mathfrak{S}_n)$ has already been proposed in [23]. It is a first breakthrough to deal with singularities of probability distributions on rankings, however it entirely relies on the algebraic structure of $\mathfrak{S}_n$. It may be thus viewed as a refinement of harmonic analysis for full or bucket rankings, but does not apply efficiently to the analysis of incomplete rankings. Several approaches have been proposed to generalize the construction of multiresolution analysis and wavelet bases on discrete spaces, mostly on trees and graphs, see for instance [4], [10], [11], [34] and [38]. None of them leads however to the construction for incomplete rankings we promote in this paper, which crucially relies on the topological properties of the complex of injective words.

The use of topological tools to analyze ranked data has been introduced in [20] and then pursued in several contributions such as in [5] or [31]. Their approach consists in modeling a collection of pairwise comparisons as an oriented flow on the graph with vertices $[\![n]\!]$ where two items are linked if the pair appears at least once in the comparisons. They show that this flow admits a "Hodge decomposition" in the sense that it can be decomposed as the sum of three components, a "gradient flow" that corresponds to globally consistent rankings, a "curl flow" that corresponds to locally inconsistent rankings, and a "harmonic flow", that corresponds to globally inconsistent but locally consistent rankings. Our construction also relies on results from topology but it decomposes the information in a quite different manner, and is tailored to the situation where incomplete rankings can be of any size.

1.3 Outline of the paper

The article is structured as follows. In section 2, the mathematical formalism that gives a rigorous definition for the concept of information localization is introduced. It is explained how group-based harmonic analysis fits in this framework and why it is not adapted to localize information related to incomplete rankings, and the analysis of the latter is formulated in the setting of injective words. Section 3 contains our major contribution: the spaces W_A are constructed and the multiresolution decomposition of $L(\mathfrak{S}_n)$ in function of these spaces is exhibited. These results are interpreted in terms of multiresolution analysis and the connection with group-based harmonic analysis is thoroughly discussed. In section 4, we construct an explicit wavelet basis adapted to the multiresolution decomposition thus built. We establish its main properties and investigate its mathematical structure. Some concluding remarks are collected in section 5, where several lines of further research are also sketched. Technical proofs are deferred to the Appendix section.

2 Information localization

It is the purpose of this section to define concepts which the subsequent analysis fully rests on, while giving insights into the relevance of our construction.

2.1 Notations

Here an throughout the article, the inclusion between two sets is denoted by $\subset$, the strict inclusion by $\subsetneq$ and the disjoint union by $\sqcup$. Given a finite set E, denote by $L(E) = \{f : E \to \mathbb{R}\}$ the $|E|$-dimensional Euclidean space of real valued functions on E equipped with the canonical inner product defined by $\langle f, g\rangle = \sum_{x\in E} f(x)g(x)$ for any $(f, g) \in L(E)^2$. We denote by δ_x the Dirac

function at any point $x \in E$ and by $\mathbb{1}_S$ the indicator function of any $S \subset E$. A partition of E is a collection of nonempty pairwise disjoint subsets $\{S_1, \ldots, S_r\}$ such that $\bigsqcup_{i=1}^{r} S_i = E$.

2.2 Localizing information through marginals

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets with $|\mathcal{Y}| \leq |\mathcal{X}|$ and Π be a mapping $\Pi : \mathcal{X} \to \mathcal{Y}$. The image of a probability distribution p on $\mathcal{X}$ by Π is denoted by p_Π. It is the probability distribution on $\mathcal{Y}$ defined by $p_\Pi(y) = \sum_{x \in \Pi^{-1}(\{y\})} p(x)$. If p is the probability distribution of a random variable X on $\mathcal{X}$, then p_Π is the probability distribution of the random variable $\Pi(X)$ on $\mathcal{Y}$, and for $y \in \mathcal{Y}$, $p_\Pi(y)$ is the probability that $\Pi(X) = y$. It is straightforward to see that the two following conditions on Π are equivalent:

1. Π is surjective on $\mathcal{Y}$ and the value of $|\Pi^{-1}(\{y\})|$ is the same for all $y \in \mathcal{Y}$.
2. The image by Π of the uniform distribution on $\mathcal{X}$ is the uniform distribution on $\mathcal{Y}$, *i.e.* if $p(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$ then $p_\Pi(y) = 1/|\mathcal{Y}|$ for all $y \in \mathcal{Y}$.

We assume that these conditions are satisfied in the sequel and call the mapping Π a "marginal transformation". For p a probability distribution on $\mathcal{X}$, we call p_Π the "marginal of p associated to Π". More generally, for any function $f \in L(\mathcal{X})$, its marginal associated to Π, denoted by $f_\Pi \in L(\mathcal{Y})$, is defined by

$$f_\Pi(y) = \sum_{x \in \Pi^{-1}(\{y\})} f(x).$$

If the function f represents a signal over the space $\mathcal{X}$ such as a probability distribution or a discrete image for example, the idea is to interpret f_Π as the degraded version of f obtained when we observe it through the transformation Π. It is "degraded" in the sense that Π being not injective in general (it can be injective only if $|\mathcal{Y}| = |\mathcal{X}|$, and in this case it is isomorphic to the identity transform), f_Π is an averaged version of f and therefore "contains less information". Without being specific about any information measure, the uniform probability distribution can be naturally interpreted as that containing no information, *i.e.* the "less localized". The assumption made on Π implies that if the original signal f on $\mathcal{X}$ contains no information, then its degraded version on $\mathcal{Y}$ still contains no information.

With a marginal transformation Π is naturally associated the marginal operator

$$\begin{aligned} M_\Pi : L(\mathcal{X}) &\to L(\mathcal{Y}) \\ f &\mapsto f_\Pi. \end{aligned}$$

Notice that M_Π is not a projection because $L(\mathcal{Y})$ is not a subspace of $L(\mathcal{X})$ (unless $\mathcal{Y} = \mathcal{X}$).

If V is a supplementary subspace of $\ker M_\Pi$ in $L(\mathcal{X})$, and $f = f_{\ker M_\Pi} + f_V$ is the corresponding decomposition of a function $f \in L(\mathcal{X})$, one has immediately $f_\Pi = M_\Pi(f) = M_\Pi(f_V)$. We can thus claim that f_V provides the same amount of information as f_Π. This means that data analysis on the space $L(\mathcal{Y})$ can be done equivalently on any supplementary space of $\ker M_\Pi$ in $L(\mathcal{X})$. The most natural choice is then surely the orthogonal supplementary $(\ker M_\Pi)^\perp$, because the latter is exactly the space of functions in $L(\mathcal{X})$ that are constant on each $\Pi^{-1}(\{y\})$ for $y \in \mathcal{Y}$ (the proof is straightforward and left to the reader).

In practice however, the signal f is observed through a finite family of marginal transformations $(\Pi_i)_{i \in I}$, and we would like to "localize" as much as possible the information related to a specific

transformation Π and not to the others. For two marginal transformations Π_1 and Π_2, we say that the subspace W_1 of $L(\mathcal{X})$ "fully localizes" the information related to Π_1 with respect to Π_2 if it satisfies the two conditions listed below:

- $W_1 \cap \ker M_{\Pi_1} = \{0\}$ (it localizes information related to Π_1),
- $W_1 \subset \ker M_{\Pi_2}$ (it localizes information that is not contained in Π_2).

Notice that there is no reason that $(\ker M_{\Pi_1})^\perp$ satisfies the latter condition for any marginal transformation Π_2.

In the general case, the definition of a space that localizes the information related to a marginal transformation Π with respect to the others depends on the relations between all the transformations of the considered family. One particularly important relation is the *refinement*. We say that the transformation Π_2 is a refinement of Π_1 if $\ker M_{\Pi_2} \subset \ker M_{\Pi_1}$. In that case, there exists a surjective linear mapping from the image of Π_2 to the image of Π_1, and we can say that Π_1 degrades the information more than Π_2 in the sense that the information related to Π_1 can be recovered from the marginal associated to Π_2 (through this surjective linear mapping) whereas the opposite is not true.

2.3 Group-based harmonic analysis on $\mathfrak{S}_n$

When the original signal space $\mathcal{X}$ is a finite group G, we can consider marginal transformations defined through its actions (see the Appendix section for some background in group theory). Let $\mathcal{Y}$ be a finite set on which G acts transitively, by $(g, y) \mapsto g \cdot y$. To each $y_0 \in \mathcal{Y}$, we associate the marginal transformation

$$\begin{aligned} \Pi_{y_0} : G &\to \mathcal{Y} \\ g &\mapsto g \cdot y_0. \end{aligned}$$

It satisfies the two conditions of a marginal transformation. First, Π_{y_0} is surjective on $\mathcal{Y}$ because the action of G on $\mathcal{Y}$ is transitive. Second, for $y \in \mathcal{Y}$, the set $\Pi_{y_0}^{-1}(\{y\}) = \{g \in G \mid g \cdot y_0 = y\}$ is a left coset of the stabilizer of y_0, $\{g \in G \mid g \cdot y_0 = y_0\}$, and thus have same cardinality. The interpretation behind this marginal transformation is as follows. If p is a probability distribution on G, then $p(g)$ is the probability of drawing the element g in G and $p_{\Pi_{y_0}}(y)$ is the probability of drawing an element $g \in G$ such that $g \cdot y_0 = y$. For a function $f \in L(G)$, the collection of all its marginals associated to the transformations Π_{y_0} for $y_0 \in \mathcal{Y}$ can be gathered in the $|\mathcal{Y}|$-squared matrix $T_{\mathcal{Y}}(f)$ defined by

$$[T_{\mathcal{Y}}(f)]_{y,y_0} = f_{\Pi_{y_0}}(y),$$

each column representing a marginal. Now, by linearity, $T_{\mathcal{Y}}(f) = \sum_{g \in G} f(g) T_{\mathcal{Y}}(\delta_g)$ for all $f \in L(G)$, and it is easy to see that for $g \in G$, $T_{\mathcal{Y}}(\delta_g)$ is actually the translation operator on $L(\mathcal{Y})$, *i.e.* for all $F \in L(\mathcal{Y})$ and $y \in \mathcal{Y}$,

$$\left(T_{\mathcal{Y}}(\delta_g) F\right)(y) = F(g^{-1} \cdot y).$$

In other words, $g \mapsto T_{\mathcal{Y}}(\delta_g)$ is the permutation representation of G on $L(\mathcal{Y})$ associated to the action $(g, y) \mapsto g \cdot y$. This means that the information contained in the collection of marginals $(\Pi_{y_0})_{y_0 \in \mathcal{Y}}$ can be decomposed using group representation theory, which is the general principle of harmonic analysis. Harmonic analysis on a finite group G is defined as the decomposition of $L(G)$ into irreducible representations of G, see [6]. These components are invariant under translation

and each localize the information related to a specific "frequency". The symmetric group has an additional particularity: each of its irreducible representations is "associated to" one specific meaningful permutation representation and thus allows to localize the information of the associated marginals. Let us develop this interpretation.

The symmetric group $\mathfrak{S}_n$ is the set of all the bijective mappings $\sigma : [\![n]\!] \to [\![n]\!]$ equipped with the composition law $(\sigma, \tau) \mapsto \sigma\tau$ defined by $\sigma\tau(i) = \sigma(\tau(i))$ for $i \in [\![n]\!]$. Irreducible representations of $\mathfrak{S}_n$, are indexed by partitions of n, *i.e.* tuples $\lambda = (\lambda_1, \dots, \lambda_r) \in [\![n]\!]^r$ such that $\lambda_1 \geq \cdots \geq \lambda_r$ and $\lambda_1 + \cdots + \lambda_r = n$, with $r \in \{1, \dots n\}$. The irreducible representation indexed by λ is denoted by S^λ, and its dimension by d_λ. The harmonic decomposition of $L(\mathfrak{S}_n)$ thus writes

$$L(\mathfrak{S}_n) \cong \bigoplus_{\lambda \vdash n} d_\lambda S^\lambda,$$

where $\lambda \vdash n$ means that the sum is taken on all the partitions λ of n (see the Appendix section). The spaces S^λ are called the Specht modules. For $\mathcal{B} = (B_1, \dots, B_r)$ an ordered partition of $[\![n]\!]$ and $\sigma \in \mathfrak{S}_n$, we set

$$\sigma \cdot \mathcal{B} = (\sigma(B_1), \dots, \sigma(B_r)),$$

where $\sigma(B) = \{\sigma(b) \mid b \in B\}$, for $B \subset [\![n]\!]$. This defines an action of $\mathfrak{S}_n$ on the set of all ordered partitions of $[\![n]\!]$. The shape of an ordered partition of $[\![n]\!]$ $\mathcal{B} = (B_1, \dots, B_r)$ is the tuple $(|B_1|, \dots, |B_r|)$. It is easy to see that the orbits of $\mathfrak{S}_n$ are the sets of ordered partitions of n of a given shape. For $\lambda \vdash n$, we denote by $\mathrm{Part}_\lambda([\![n]\!])$ be the set of ordered partitions of $[\![n]\!]$ of shape λ, and define $M^\lambda = L(\mathrm{Part}_\lambda([\![n]\!]))$, called a Young module. In this context, the marginal transformation associated to a $\mathcal{B}_0 \in \mathrm{Part}_\lambda([\![n]\!])$ is defined by

$$\begin{aligned} \Pi_{\mathcal{B}_0} : \mathfrak{S}_n &\to \mathrm{Part}_\lambda([\![n]\!]) \\ \sigma &\mapsto \sigma \cdot \mathcal{B}_0. \end{aligned}$$

The marginal of a function $f \in L(\mathfrak{S}_n)$ associated to this transformation is denoted by $f_{\mathcal{B}_0}$ and is called a λ-marginal. The $|\mathrm{Part}_\lambda([\![n]\!])|$-square matrix $T_\lambda(f)$ that gathers all the λ-marginals of f is equal to the sum $\sum_{\sigma \in \mathfrak{S}_n} f(\sigma) T_\lambda(\delta_\sigma)$, where $T_\lambda(\delta_\sigma)$ is the matrix of the permutation representation of $\mathfrak{S}_n$ on M^λ taken in σ. Now it happens that the Specht module S^λ can be defined as a subspace of M^λ, see [6]. This means that the information localized by S^λ in the harmonic decomposition of $L(\mathfrak{S}_n)$ is contained in the λ-marginals. This information is also specific to λ-marginals in a certain way.

The dominance order on partitions of n is the partial order defined, for $\lambda = (\lambda_1, \dots, \lambda_r)$ and $\mu = (\mu_1, \dots, \mu_s)$ by $\lambda \trianglerighteq \mu$ if for all $j \in \{1, \dots, r\}$, $\sum_{i=1}^j \lambda_i \geq \sum_{i=1}^j \mu_i$. When $\lambda \trianglerighteq \mu$ and $\lambda \neq \mu$ we write $\lambda \triangleright \mu$. The decomposition of the Young module M^λ for $\lambda \vdash n$ is given by Young's rule (see [6]):

$$M^\lambda \cong \bigoplus_{\mu \vdash n} K_{\mu,\lambda} S^\mu, \qquad \text{where} \quad \begin{cases} K_{\mu,\lambda} = 0 & \text{if } \mu \trianglelefteq \lambda \\ K_{\lambda,\lambda} = 1 & \\ K_{\mu,\lambda} \geq 1 & \text{if } \mu \triangleright \lambda \end{cases},$$

i.e.

$$M^\lambda \cong S^\lambda \oplus \bigoplus_{\mu \triangleright \lambda} K_{\mu,\lambda} S^\mu.$$

This means that for a given $\lambda \vdash n$, the Specht module S^λ localizes the information of M^λ that is not contained in the M^μ's for $\mu \triangleright \lambda$. In this sense, S^λ contains the information that is specific to λ-marginals and not the others.

2.4 "Absolute" and "relative" rank information

The precedent subsection shows that harmonic analysis on $\mathfrak{S}_n$ consists in localizing information specific to collections of marginal transformations $(\Pi_{\mathcal{B}})_{\mathcal{B}\in \mathrm{Part}_\lambda([\![n]\!])}$, for $\lambda \vdash n$. Let p be a probability distribution on $\mathfrak{S}_n$ and Σ a random permutation of law p. If $S \subset [\![n]\!]$ represents an event $\mathcal{E}$ on the random variable X, we denote by $\mathbb{P}[\mathcal{E}]$ the probability of this event, *i.e.* $\mathbb{P}[\mathcal{E}] = \sum_{\sigma\in S} p(\sigma)$. For $\lambda = (\lambda_1, \ldots, \lambda_r) \vdash n$ and $\mathcal{B} = (B_1, \ldots, B_r) \in \mathrm{Part}_\lambda([\![n]\!])$, we have for any $\mathcal{B}' = (B'_1, \ldots, B'_r) \in \mathrm{Part}_\lambda([\![n]\!])$,

$$p_{\mathcal{B}}(\mathcal{B}') = \mathbb{P}\left[\Sigma(B_1) = B'_1, \ldots, \Sigma(B_r) = B'_r\right].$$

To gain more insight into the interpretation of these marginals, let us consider first the simple case $\lambda = (n-1, 1)$. Ordered partitions of $[\![n]\!]$ of shape $(n-1, 1)$ are necessarily of the form $([\![n]\!]\setminus\{i\}, \{i\})$, with $i \in [\![n]\!]$. Then for $(i, j) \in [\![n]\!]^2$, we have the simplification

$$\mathbb{P}\left[\Sigma([\![n]\!] \setminus \{i\}) = [\![n]\!] \setminus \{j\},\ \Sigma(\{i\}) = \{j\}\right] = \mathbb{P}\left[\Sigma(i) = j\right].$$

The marginal of p associated to $([\![n]\!]\setminus\{i\}, \{i\})$ is thus the probability distribution $(\mathbb{P}[\Sigma(i) = j])_{j\in[\![n]\!]}$ on $[\![n]\!]$. From a ranking point of view, this is the law of the rank of item i. The matrix $T_{(n-1,1)}(p)$ that gathers all the $(n-1, 1)$-marginals of p is given by

$$T_{(n-1,1)}(p) = \begin{pmatrix} \mathbb{P}[\Sigma(1) = 1] & \cdots & \mathbb{P}[\Sigma(n) = 1] \\ \vdots & \ddots & \vdots \\ \mathbb{P}[\Sigma(1) = n] & \cdots & \mathbb{P}[\Sigma(n) = n] \end{pmatrix},$$

where the marginal of p associated to $([\![n]\!] \setminus \{i\}, \{i\})$ is represented by column i. It is easy to see that this matrix is bistochastic, and that the row i represents the probability distribution $(\mathbb{P}[\Sigma^{-1}(i) = j])_{j\in[\![n]\!]}$ on $[\![n]\!]$. This is the probability distribution of the index of the element ranked at the i^{th} position. In both cases, the distribution captures the information about an "absolute rank", in the sense that it is the rank of an item inside a ranking implying all the n items. Either we consider the distribution of the absolute ranks of a fixed item i, or else we consider the distribution of the item having a fixed absolute rank i.

More generally for $k \in \{1, \ldots, n-1\}$, elements of $\mathrm{Part}_{(n-k,k)}([\![n]\!])$ are of the form $([\![n]\!] \setminus A, A)$ with $A \in \binom{[\![n]\!]}{k}$, and the marginal law of p associated to $([\![n]\!] \setminus A, A)$ is the probability distribution $(\mathbb{P}[\Sigma(A) = B])_{|B|=k}$ on $\binom{[\![n]\!]}{k}$. From a ranking point of view, $\mathbb{P}[\Sigma(A) = B]$ is the probability that the items of A are ranked at the absolute positions of B, regardless of their order inside these positions. In the general case, for $\lambda = (\lambda_1, \ldots, \lambda_r) \vdash n$, $\mathcal{B} = (B_1, \ldots, B_r) \in \mathrm{Part}_\lambda([\![n]\!])$ and $\mathcal{B}' = (B'_1, \ldots, B'_r) \in \mathrm{Part}_\lambda([\![n]\!])$, $\mathbb{P}[\Sigma \cdot \mathcal{B} = \mathcal{B}']$ is the probability that the items of B_i are ranked at the absolute positions of B'_i, for $i \in \{1, \ldots, r\}$.

Example 1. We give an illustration of this type of marginals on a real dataset with $n = 4$, studied in [7]. It is composed of 2262 answers of German citizens who were asked to rank the desirability of four political goals, that we consider as items 1, 2, 3 and 4. Each ranking of these four items received a certain number of votes. Normalizing by 2262, the total number of votes,we obtain a probability distribution p on $\mathfrak{S}_4$. It is represented in figure 1, where the x-axis represents the 24 elements of $\mathfrak{S}_4$, denoted by $a_1a_2a_3a_4$ instead of $a_1 \prec a_2 \prec a_3 \prec a_4$ and ordered by the lexicographic order, *i.e.* $1234, 1243, ..., 4321$.

There are 5 partitions of 4: $(4), (3, 1), (2, 2), (2, 1, 1), (1, 1, 1, 1)$. There is only one (4)-marginal, it is the constant $\sum_{\sigma\in\mathfrak{S}_4} p(\sigma) = 1$, and there are 24 $(1, 1, 1, 1)$-marginals, the translations of p.

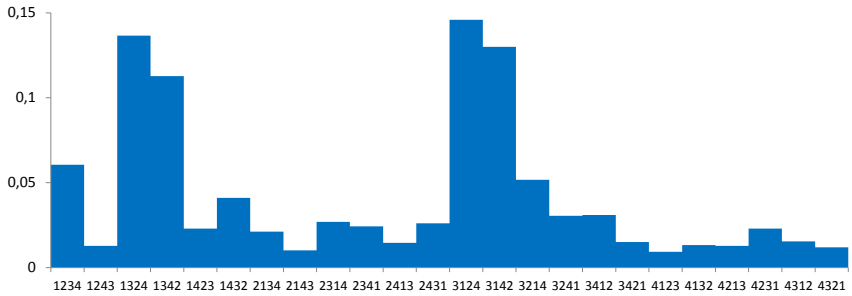


Figure 1: Probability distribution p of the rankings of the four political goals

We consider the marginals of the three other types. Let Σ be a random permutation of law p. The $(3,1)$ marginals are the laws of the random variables $\Sigma(i)$, representing the rank of item i, for $i \in [\![4]\!]$. The $(2,2)$-marginals are the laws of the random variables $\{\Sigma(i), \Sigma(j)\}$ and the $(2,1,1)$ are the laws of the random variables $(\Sigma(i), \Sigma(j))$, for $(i,j) \in [\![n]\!]^2$ with $i \neq j$. All these marginals are represented in figure 2 (with different scales).

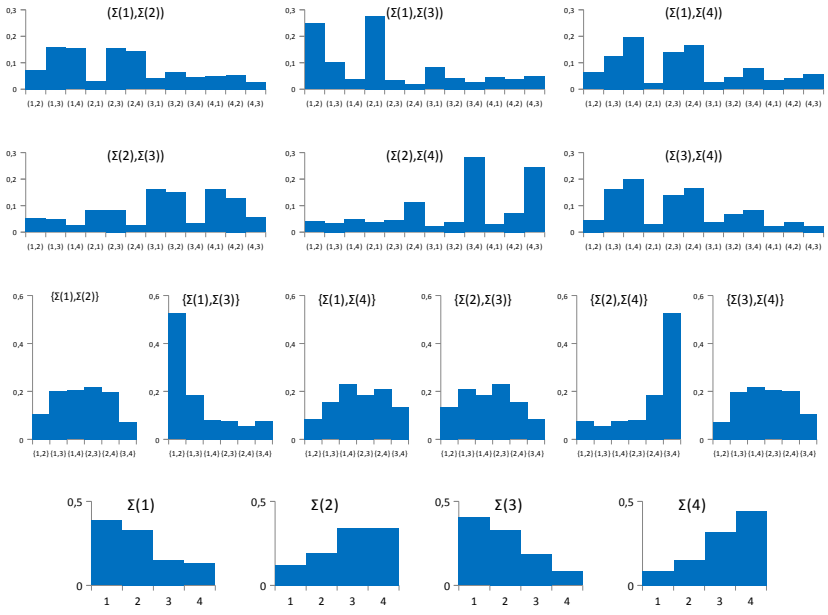


Figure 2: Absolute λ-marginals of p, for $\lambda = (3,1), (2,2), (2,1,1)$

In the analysis of incomplete rankings, we are not interested in absolute rank information, but in "relative" rank information. When incomplete rankings are observed on a subset of items $A \in \mathcal{P}([\![n]\!])$, the information we have access to is about the ranks of the items of A relatively to A. In the same way, the prediction of a ranking on A only involves the information about the ranks

of the items of A relatively to A. This is the fundamental difference between the analysis of full rankings or bucket orders (also called partial rankings) and the analysis of incomplete rankings. This implies that the marginal transformations and the information localization involved are completely different. Let Σ be a random permutation of law p on $\mathfrak{S}_n$. In the analysis of incomplete rankings, we are interested in probabilities of the form

$$\mathbb{P}[\Sigma(a_1) < \cdots < \Sigma(a_k)],$$

for $k \in \{2, \ldots, n\}$ and $\{a_1, \ldots, a_k\} \subset [\![n]\!]$. So we are not interested in the values $\Sigma(a_1), \ldots, \Sigma(a_k)$, but only in their ordering, which induce a ranking of the items $a_1, \ldots, a_k$. We are thus interested in the marginals p_A of p defined in the Introduction section, for $A \in \mathcal{P}([\![n]\!])$.

Example 2. Considering the same example as before, we represent all the marginals of p involved in the analysis of incomplete rankings. For each $A \in \mathcal{P}([\![4]\!])$, the marginal p_A is represented in figure 3 by a graph with the x-axis constituted of the elements of $\mathfrak{S}'_A$ ordered by the lexicographic order.

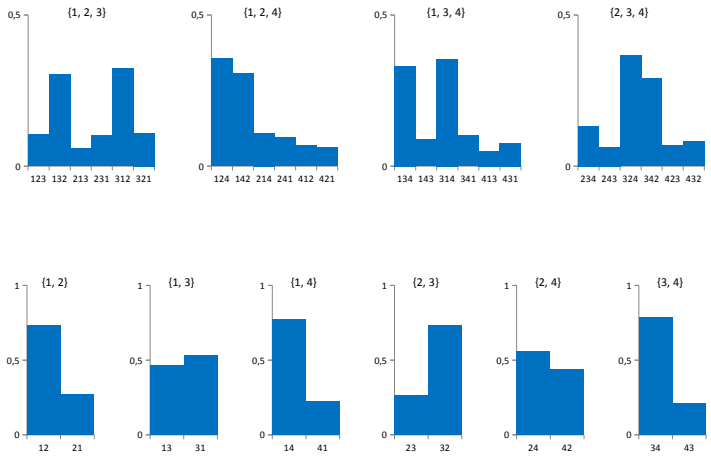


Figure 3: Relative marginals of p on subsets $A \subset [\![4]\!]$ with $|A| = 2$ or 3

It is obvious that each of the two families of marginal transformations leads to the analysis of completely different objects: full rankings or bucket orders involving absolute rank information, or incomplete rankings involving relative rank information. There is a way to handle incomplete rankings with $\mathfrak{S}_n$-based harmonic analysis, as it is done in [22], but it is not really adapted and it does not provide a powerful general framework. This can be achieved only by fully exploiting the structure of incomplete rankings and considering the right marginal transformations.

In this case, the marginal transformations suited for the analysis of incomplete rankings map a permutation σ to the ranking it induces on a subset of items $A = \{a_1, \ldots, a_k\} \in \mathcal{P}([\![n]\!])$, through the order of the values $\sigma(a_1), \ldots, \sigma(a_k)$. This definition is however not easy to use and thus not convenient to characterize the structure of incomplete rankings. It happens that they can be defined from another point of view that fits with the mathematical structure of incomplete rankings. It

comes from the observation that the ranking induced by a full ranking on a subset of items A is obtained by keeping only the items of A, in the same order. More specifically, if σ corresponds to the full ranking $a_1 \prec \cdots \prec a_n$ on $[\![n]\!]$, the ranking it induces on A is given by $a_{i_1} \prec \cdots \prec a_{i_{|A|}}$ where $i_1 < \cdots < i_{|A|}$ and $A = \{a_{i_1}, \ldots, a_{i_{|A|}}\}$. This perspective is best expressed in the language of injective words.

2.5 Analysis of incomplete rankings through injective words

An injective word over $[\![n]\!]$ is an expression $\omega = \omega_1 \ldots \omega_k$ where $1 \leq k \leq n$ and $\omega_1, \ldots, \omega_k$ are distinct elements of $[\![n]\!]$. The content of the word $\omega = \omega_1 \ldots \omega_k$ is $c(\omega) = \{\omega_1, \ldots, \omega_k\}$, and its size is $|\omega| := |c(\omega)|$. The empty word $\overline{0}$ is by convention the unique word of size 0 and content $\emptyset$. A subword of a word $\omega = \omega_1 \ldots \omega_k \in \Gamma_n$ is an expression $\omega_{i_1} \ldots \omega_{i_r}$ with $1 \leq i_1 < \cdots < i_r \leq k$. We denote by Γ_n the set of injective words over $[\![n]\!]$ and for $A \subset [\![n]\!]$ and $k \in \{0, \ldots, n\}$, we set $\Gamma(A) = \{\omega \in \Gamma_n \mid c(\omega) = A\}$ and $\Gamma^k = \{\omega \in \Gamma_n \mid |\omega| = k\}$. We thus have

$$\Gamma_n = \bigsqcup_{k=0}^{n} \Gamma^k = \bigsqcup_{k=0}^{n} \bigsqcup_{|A|=k} \Gamma(A). \tag{3}$$

To each incomplete ranking $\pi = a_1 \prec \cdots \prec a_k$, we associate the corresponding injective word $a_1 \ldots a_k$, and we still denote it by π. The sets $\mathfrak{S}'_A$ and $\Gamma(A)$ are thus identified for $A \in \mathcal{P}([\![n]\!])$, in particular $\mathfrak{S}_n$ is identified to $\Gamma([\![n]\!])$, and both interpretations will be used indifferently in the sequel.

The language of injective words has two major advantages for the analysis of incomplete rankings. The first is that it is well suited to express the marginal transformations that we want to consider and their properties. Let $(A, B) \in \mathcal{P}([\![n]\!])^2$ with $A \subset B$ and $\sigma = a_1 \ldots a_{|B|} \in \mathfrak{S}'_B$ representing the ranking $a_1 \prec \cdots \prec a_{|B|}$. Then the ranking induced by σ on A is represented by the unique subword of σ with content equal to A. The latter is obtained by deleting from $a_1 \ldots a_{|B|}$ all the a_i's that do not belong to A. We denote by $\sigma_{|A}$ the ranking induced by σ on A as well as the injective word representing it. The marginal transformations of interest in the analysis of incomplete rankings are thus defined by

$$\begin{aligned} \Pi_A : \mathfrak{S}_n &\to \mathfrak{S}'_A \\ \sigma &\mapsto \sigma_{|A}, \end{aligned}$$

for $A \in \mathcal{P}([\![n]\!])$. We denote respectively by f_A and M_A the marginal of a function $f \in (\mathfrak{S}_n)$ and the marginal operator associated to Π_A (these notations are the same as in the introduction). Recall that for $\pi \in \mathfrak{S}'_A$ viewed as a mapping $A \to \{1, \ldots, |A|\}$, the set $\mathfrak{S}_n(\pi)$ is defined as $\mathfrak{S}_n(\pi) = \{\sigma \in \mathfrak{S}_n \mid \text{for all } (a, b) \in A^2 \text{ such that } \pi(a) < \pi(b),\ \sigma(a) < \sigma(b)\}$. Viewing now $\pi \in \mathfrak{S}'_A$ as an injective word, it is clear that $\mathfrak{S}_n(\pi) = \{\sigma \in \mathfrak{S}_n \mid \sigma_{|A} = \pi\} = \Pi^{-1}(\{\pi\})$. More generally we define, for $(A, B) \in \mathcal{P}([\![n]\!])^2$ with $A \subset B$ and $\pi \in \mathfrak{S}'_A$, $\mathfrak{S}'_B(\pi) = \{\sigma \in \mathfrak{S}'_B \mid \sigma_{|A} = \pi\}$, with $\mathfrak{S}_n(\pi) := \mathfrak{S}'_{[\![n]\!]}(\pi)$. The fact that Π_A is a marginal transformation is thus a direct consequence of the following lemma (for $B = [\![n]\!]$), its technical proof is postponed to the Appendix section.

Lemma 1. *Let $(A, B) \in \mathcal{P}([\![n]\!])^2$ with $A \subset B$. Then $\{\mathfrak{S}'_B(\pi)\}_{\pi \in \mathfrak{S}'_A}$ is a partition of $\mathfrak{S}'_B$ and for all $\pi \in \mathfrak{S}'_A$, $|\mathfrak{S}'_B(\pi)| = |B|!/|A|!$.*

The refinement relations inside the family of marginal transformations $(\Pi_A)_{A \in \mathcal{P}([\![n]\!])}$ rely on the structure of injective words. For $\pi \in \Gamma_n$ with $|\pi| < n$, $b \in [\![n]\!] \setminus c(\pi)$ and $i \in \{1, \ldots, |\pi| + 1\}$, we

denote by $\pi \triangleleft_i b$ the word obtained by inserting b in i^{th} position in π. The following lemma, the proof of which is straightforward, is the base of the refinement relations.

Lemma 2. *Let $A \subsetneq B \subset [\![n]\!]$ and $\pi \in \mathfrak{S}'_A$. For all $b \in B \setminus A$, $\mathfrak{S}'_B(\pi) = \bigsqcup_{i=1}^{|A|+1} \mathfrak{S}'_B(\pi \triangleleft_i b)$. In particular, $\mathfrak{S}'_{A\cup\{b\}}(\pi) = \{\pi \triangleleft_1 b, \dots, \pi \triangleleft_{|A|+1} b\}$.*

Proposition 1. *For $A, B \in \mathcal{P}([\![n]\!])$ with $A \subset B$, Π_B is a refinement of Π_A, i.e.*

$$\ker M_B \subset \ker M_A.$$

Proof. Let $A \in \mathcal{P}([\![n]\!])$ and $b \in [\![n]\!] \setminus A$. For $f \in L(\mathfrak{S}_n)$, lemma 2 gives, for all $\pi \in \mathfrak{S}'_A$,

$$M_A f(\pi) = \sum_{\sigma \in \mathfrak{S}_n(\pi)} f(\sigma) = \sum_{i=1}^{|A|+1} \sum_{\sigma \in \mathfrak{S}_n(\pi \triangleleft_i b)} f(\sigma) = \sum_{i=1}^{|A|+1} M_{A\cup\{b\}} f(\pi \triangleleft_i b).$$

This implies that $\ker M_{A\cup\{b\}} \subset \ker M_A$ and the proof is concluded by induction. □

The second major advantage of the language of injective words is that it allows to define a global framework for all incomplete rankings. To this purpose, we see the elements of $L(\Gamma_n)$ as free linear combinations of injective words, also called chains, *i.e.* expressions of the form $x = \sum_{\omega \in \Gamma_n} x(\omega)\omega$, where ω refers at the same time to a word in Γ_n and to the Dirac function of this word in $L(\Gamma_n)$. Notice then that $\overline{0}$ denotes the Dirac function in the empty word, whereas 0 denotes the function equal to 0 for all $\omega \in \Gamma_n$, and that the indicator function of a set $S \subset \Gamma_n$ is equal to the sum of the Dirac functions in its elements

$$\mathbb{1}_S = \sum_{\sigma \in S} \sigma.$$

By definition, the marginal operator M_A applied to the Dirac function of $\sigma \in \mathfrak{S}_n$ in $L(\mathfrak{S}_n)$ is equal to the Dirac function of $\sigma_{|A}$ in $L(\mathfrak{S}'_A)$. Using the chain notation, this gives:

$$M_A \sigma = \sigma_{|A}. \tag{4}$$

A function in $L(\Gamma(A))$ for $A \subset [\![n]\!]$ is thus directly seen as a chain in $L(\Gamma_n)$, and by equation (3), we have $L(\Gamma_n) = \bigoplus_{k=0}^{n} \bigoplus_{|A|=k} L(\Gamma(A))$. This decomposition allows to embed $L(\mathfrak{S}_n)$, all the spaces of marginals $L(\mathfrak{S}'_A)$ for $A \in \mathcal{P}([\![n]\!])$ and the spaces $L(\Gamma(A))$ for $|A| \leq 1$ into one general space, that is $L(\Gamma_n)$. For $n = 4$, $L(\Gamma_4)$ decomposes as follows.

$$L\left(\mathfrak{S}_4\right)$$
$$L\left(\mathfrak{S}'_{\{1,2,3\}}\right) \oplus L\left(\mathfrak{S}'_{\{1,2,4\}}\right) \oplus L\left(\mathfrak{S}'_{\{1,3,4\}}\right) \oplus L\left(\mathfrak{S}'_{\{2,3,4\}}\right)$$
$$L\left(\mathfrak{S}'_{\{1,2\}}\right) \oplus L\left(\mathfrak{S}'_{\{1,3\}}\right) \oplus L\left(\mathfrak{S}'_{\{1,4\}}\right) \oplus L\left(\mathfrak{S}'_{\{2,3\}}\right) \oplus L\left(\mathfrak{S}'_{\{2,4\}}\right) \oplus L\left(\mathfrak{S}'_{\{3,4\}}\right)$$
$$L(\Gamma(\{1\})) \oplus L(\Gamma(\{2\})) \oplus L(\Gamma(\{3\})) \oplus L(\Gamma(\{4\}))$$
$$L(\Gamma(\overline{0}))$$

This embedding allows to model all possible observations of incomplete rankings. Indeed, let $\mathcal{A} \subset \mathcal{P}([\![n]\!])$ be an observation design. Then for each $A \in \mathcal{A}$, the variability of the observed rankings on A is represented by a probability distribution $P_A \in L(\mathfrak{S}'_A)$. The total variability of the observed rankings is thus represented by the collection $(P_A)_{A\in\mathcal{A}} \in \bigoplus_{A\in\mathcal{A}} L(\mathfrak{S}'_A) \subset L(\mathfrak{S}_n)$.

Example 3. Let us assume that we observe incomplete rankings on $[\![4]\!]$ through the observation design $\mathcal{A} = \{\{1,3\},\{2,4\},\{3,4\},\{1,2,3\},\{1,3,4\}\}$. Then the collection of probability distributions is an element of the sum of the spaces in bold, in the following representation.

$$L\left(\mathfrak{S}_4\right)$$
$$\mathbf{L}\left(\mathfrak{S}'_{\mathbf{\{1,2,3\}}}\right) \oplus L\left(\mathfrak{S}'_{\{1,2,4\}}\right) \oplus \mathbf{L}\left(\mathfrak{S}'_{\mathbf{\{1,3,4\}}}\right) \oplus L\left(\mathfrak{S}'_{\{2,3,4\}}\right)$$
$$L\left(\mathfrak{S}'_{\{1,2\}}\right) \oplus \mathbf{L}\left(\mathfrak{S}'_{\mathbf{\{1,3\}}}\right) \oplus L\left(\mathfrak{S}'_{\{1,4\}}\right) \oplus L\left(\mathfrak{S}'_{\{2,3\}}\right) \oplus \mathbf{L}\left(\mathfrak{S}'_{\mathbf{\{2,4\}}}\right) \oplus \mathbf{L}\left(\mathfrak{S}'_{\mathbf{\{3,4\}}}\right)$$
$$L(\Gamma(\{1\})) \oplus L(\Gamma(\{2\})) \oplus L(\Gamma(\{3\})) \oplus L(\Gamma(\{4\}))$$
$$L(\Gamma(\overline{0}))$$

Notice however that we are not interested in performing data analysis in the space $\bigoplus_{A\in\mathcal{A}} L(\mathfrak{S}'_A)$ but in its subspace $\mathbb{M}_\mathcal{A}$ of the collections $(f_A)_{A\in\mathcal{A}} \in \bigoplus_{A\in\mathcal{A}} L(\mathfrak{S}'_A)$ that satisfy condition $(*)$, $\mathbb{M}_\mathcal{A} = M_\mathcal{A}(L(\mathfrak{S}_n))$. This embedding remains nonetheless very convenient to define global operators that exploit the structure of injective words.

Definition 1 (Deletion operator)**.** Let $a \in [\![n]\!]$. For $\pi \in \Gamma_n$ such that $a \in c(\pi)$, we denote by $\pi \setminus \{a\}$ the word obtained by deleting the letter a in the word π. We extend this operation into the operator $\varrho_a : L(\Gamma_n) \to L(\Gamma_n)$, defined on a Dirac function π by

$$\varrho_a\pi = \begin{cases} \pi \setminus \{a\} & \text{if } a \in c(\pi) \\ \pi & \text{otherwise.} \end{cases}$$

For $a_1, a_2 \in [\![n]\!]$, it is obvious that $\varrho_{a_1}\varrho_{a_2} = \varrho_{a_2}\varrho_{a_1}$. This allows to define, for $A = \{a_1, \ldots, a_k\} \subset [\![n]\!]$, $\varrho_A = \varrho_{a_1} \ldots \varrho_{a_k}$. We set by convention $\varrho_\emptyset x = x$ for all $x \in L(\Gamma_n)$.

Remark 1. Notice that for any $\pi \in \Gamma_n$, $\varrho_{c(\pi)}\pi = \overline{0}$. This implies that for $A \subset [\![n]\!]$ and $x \in L(\Gamma(A))$, $\varrho_A x = \left[\sum_{\pi\in\Gamma(A)} x(\pi)\right] \overline{0}$.

The family of spaces $(L(\Gamma(A)))_{A\subset[\![n]\!]}$ equipped with the family of operators $(\varrho_{B\setminus A})_{A\subset B\subset[\![n]\!]}$ is a projective system, *i.e.* for all $A \subset B \subset C \subset [\![n]\!]$,

- $\varrho_{B\setminus A} : L(\Gamma(B)) \to L(\Gamma(A))$,
- $\varrho_{A\setminus A} x = x \qquad$ for all $x \in L(\Gamma(A))$,
- $\varrho_{B\setminus A}\varrho_{C\setminus B} = \varrho_{C\setminus A}$.

It is represented for $n = 4$ in figure 4.

With these notations, assumption $(*)$ for a family $(f_A)_{A\in\mathcal{P}([\![n]\!])}$ becomes: for $A \in \mathcal{P}([\![n]\!])$ with $|A| < n$ and $b \in [\![n]\!] \setminus A$,

$$f_A(\pi) = \sum_{i=1}^{|A|+1} f_{A\cup\{b\}}(\pi \lhd_i b) = \varrho_b f_{A\cup\{b\}}(\pi),$$

for all $\pi \in \mathfrak{S}'_A$. The projective system properties then imply more generally that for any $(A, B) \in \mathcal{P}([\![n]\!])$ with $A \subset B$,

$$\varrho_{B\setminus A} f_B = f_A.$$

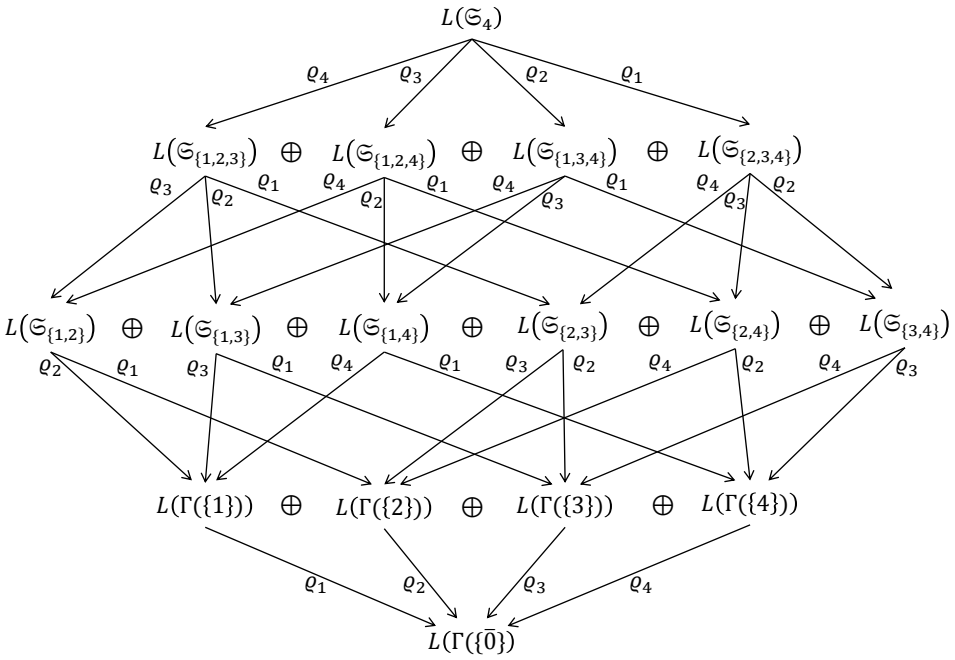


Figure 4: Projective system $((L(\Gamma(A)))_{A\subset[\![n]\!]}, (\varrho_{B\setminus A})_{A\subset B\subset[\![n]\!]})$ for $n = 4$

Example 4. We keep the same example as before: the number of items is $n = 4$ and the observation design is $\mathcal{A} = \{\{1,3\},\{2,4\},\{3,4\},\{1,2,3\},\{1,3,4\}\}$. The relations imposed on an element $(f_A)_{A\in\mathcal{A}} \in \mathbb{M}_{\mathcal{A}}$ are represented in figure 5.

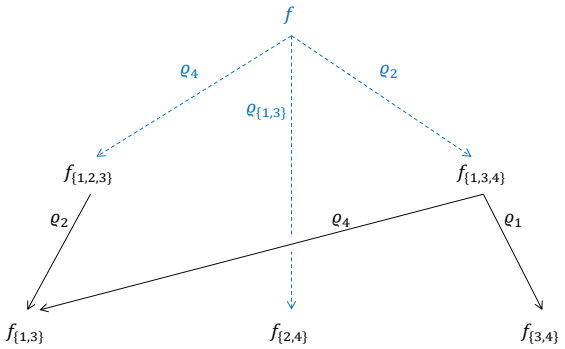


Figure 5: Projectivity conditions on $\mathbb{M}_{\mathcal{A}}$ for $\mathcal{A} = \{\{1,3\},\{2,4\},\{3,4\},\{1,2,3\},\{1,3,4\}\}$

Now, let $(A, B) \in \mathcal{P}([\![n]\!])^2$ with $A \subset B$ and $\sigma \in \mathfrak{S}'_B$. By definition, $\sigma_{|A}$ is the word obtained by deleting in σ all the elements that are not in A, so $\sigma_{|A} = \sigma \setminus (B \setminus A) = \varrho_{B\setminus A}\sigma$. In particular for $B = [\![n]\!]$, we have from equation (4)

$$M_A = \varrho_{[\![n]\!]\setminus A}. \tag{5}$$

All the marginals operators can thus be expressed in terms of deletion operators. For an element $(f_A)_{A\in\mathcal{A}}$ of a observation design $\mathcal{A} \subset \mathcal{P}([\![n]\!])$, we express each f_A as the marginal on A of a function f. Our goal is to obtain a decomposition of f into components that have a localized effect on the f_A's. More precisely, we want a decomposition of f of the form

$$f = \tilde{f}_0 + \sum_{B\in\bigcup_{A\in\mathcal{A}}\mathcal{P}(A)} \tilde{f}_B$$

such that for any $A \in \mathcal{A}$,

$$f_A = M_A\left[\tilde{f}_0 + \sum_{B\in\mathcal{P}(A)} \tilde{f}_B\right]. \tag{6}$$

Example 5. Using the same example as before, we represent the principle of the decomposition in figure 6.

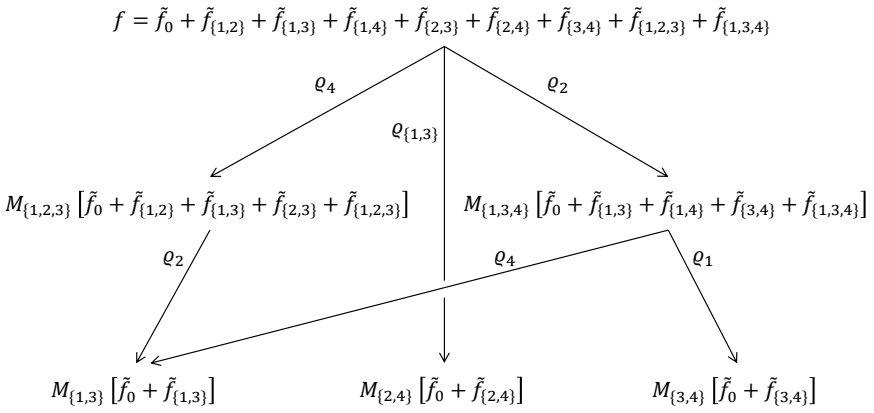


Figure 6: Decomposition of a function $f \in L(\mathfrak{S}_n)$ adapted to the observation design $\mathcal{A} = \{\{1,3\},\{2,4\},\{3,4\},\{1,2,3\},\{1,3,4\}\}$

3 The multiresolution decomposition

We now enter the details of the construction of our multiresolution decomposition of $L(\mathfrak{S}_n)$ that provides a multiresolution decomposition of $\mathbb{M}_\mathcal{A}$ for any observation design $\mathcal{A} \subset \mathcal{P}([\![n]\!])$.

3.1 Requirements for W_A

For $A \in \mathcal{P}([\![n]\!])$, we want to construct a subspace W_A of $L(\mathfrak{S}_n)$ that "localizes the information that is specific to the marginal on A and not to the others". The precise definition of this statement relies on the refinement relation on the Π_B's shown in proposition 1, namely for $B \subset B'$, $\Pi_{B'}$ is a refinement of Π_B. This implies first that W_A cannot contain the entire information related to the marginal on A, otherwise it would contain also the entire information related to the marginal on B for all $B \subset A$, which is not specific to A. So W_A cannot be a supplementary space of $\ker M_A$ but we require that all the information it localizes be contained in the marginal on A, *i.e.* that $W_A \cap \ker M_A = \{0\}$. Second, for $B \supset A$, the marginal on B contains all the information related to the marginal on A and *a fortiori* the information localized by W_A, so we have necessarily $W_A \cap \ker M_B = \{0\}$. We can require however that $W_A \subset \ker M_B$ for all $B \in \mathcal{P}([\![n]\!])$ such that $B \not\supset A$. We thus want W_A to satisfy two conditions:

1. it localizes information related to the marginal on A, *i.e.*

$$W_A \cap \ker M_A = \{0\} \tag{7}$$

2. it localizes information that is not contained in the marginals on $B \in \mathcal{P}([\![n]\!])$ for $B \not\supset A$, *i.e.*

$$W_A \subset \bigcap_{\substack{B \in \mathcal{P}([\![n]\!]) \\ B \not\supset A}} \ker M_B. \tag{8}$$

Let us first consider the case $A = [\![n]\!]$. The operator $M_{[\![n]\!]}$ is equal to the identity mapping on $L(\mathfrak{S}_n)$, so $\ker M_{[\![n]\!]} = \{0\}$ and $W_{[\![n]\!]}$ only needs to satisfy condition (8). Since we want $W_{[\![n]\!]}$ to localize all the information that is specific to $M_{[\![n]\!]}$, we define

$$W_{[\![n]\!]} = \bigcap_{\substack{B \in \mathcal{P}([\![n]\!]) \\ B \subsetneq [\![n]\!]}} \ker M_B.$$

Using proposition 1, one has $W_{[\![n]\!]} = \bigcap_{|B|=n-1} \ker M_B$. Now, if $|B| = n-1$, $[\![n]\!] \setminus B$ is necessarily of the form $\{a\}$ with $a \in [\![n]\!]$. Thus, using equation (5), we obtain

$$W_{[\![n]\!]} = \{x \in L(\mathfrak{S}_n) \mid \varrho_a(x) = 0 \text{ for all } a \in [\![n]\!]\}.$$

More generally for $A \in \mathcal{P}([\![n]\!])$, let H_A be the space

$$H_A = \{x \in L(\Gamma(A)) \mid \varrho_a(x) = 0 \text{ for all } a \in A\}. \tag{9}$$

Seeing $L(\Gamma(A))$ as the space of marginals on A, the space H_A contains, among the information related to marginals on A, the information that is specific to A and not to subsets $B \subsetneq A$. This is exactly the information that we want W_A to localize. But the elements of H_A are chains on words with content A, not $[\![n]\!]$, and H_A is not a subspace of $L(\mathfrak{S}_n)$. The space W_A must then be constructed as an embedding of H_A into $L(\mathfrak{S}_n)$. The choice of the embedding can nonetheless not be arbitrary if we want W_A to satisfy conditions (7) and (8). We are looking for a linear operator $\phi_n : L(\Gamma_n) \to L(\mathfrak{S}_n)$ such that for all $A \in \mathcal{P}([\![n]\!])$,

$$\phi_n(H_A) \cap \ker M_A = \{0\} \qquad \text{and} \qquad \phi_n(H_A) \subset \bigcap_{\substack{B \in \mathcal{P}([\![n]\!]) \\ B \not\supset A}} \ker M_B.$$

The rationale behind this is that we want ϕ_n to "pull up" all the information contained in H_A in $L(\mathfrak{S}_n)$ in a way that it does not impact the marginals on the subsets $B \in \mathcal{P}([\![n]\!])$ such that $B \not\supset A$. Mapping the Dirac function of a $\pi \in \mathfrak{S}'_A$ in $L(\Gamma_n)$ to an element of $L(\mathfrak{S}_n)$ involves necessarily the insertion of the missing items $[\![n]\!] \setminus A$. But this can be done in many different ways. In the case where $A = [\![n]\!] \setminus \{b\}$ with $b \in [\![n]\!]$, the insertion of b in an element $\pi \in \mathfrak{S}'_A$ can be done at any of the n positions. More generally for $|A| = k$, the number of ways to insert the items of $[\![n]\!] \setminus A$ in an element $\pi \in \mathfrak{S}'_A$ is equal to $n!/k!$. Perhaps the most natural embedding is to insert the items in all possible ways. The embedding operator would then be defined on the Dirac function of a word $\pi \in \Gamma^{n-1}$ with $[\![n]\!] \setminus c(\pi) = \{b\}$ by

$$\phi'_n \pi = \sum_{i=1}^{n} \pi \lhd_i b,$$

and more generally on the Dirac function of any $\pi \in \Gamma_n$ by

$$\phi'_n \pi = \mathbb{1}_{\mathfrak{S}_n(\pi)} = \sum_{\sigma \in \mathfrak{S}_n(\pi)} \sigma.$$

For $A \in \mathcal{P}([\![n]\!])$ and $\pi \in \mathfrak{S}'_A$, we have

$$M_A \phi'_n \pi = M_A \mathbb{1}_{\mathfrak{S}_n(\pi)} = \sum_{\sigma \in \mathfrak{S}_n(\pi)} \sigma_{|A} = \frac{n!}{|A|!} \pi, \quad \text{thus} \quad \phi(H_A) \cap \ker M_A = \{0\},$$

but for $B \in \mathcal{P}([\![n]\!])$ such that $B \not\subset A$ and $x \in H_A \setminus \{0\}$, $M_B \phi'_n x \neq 0$. This can be shown in the general case but it is not necessary here. We just consider a simple example to give some insights, we take $n = 3$, $A = \{1, 2\}$ and $B = \{1, 3\}$. By definition, $H_{\{1,2\}}$ is the space of chains of the form $\alpha.12 + \beta.21$ such that $\alpha + \beta = 0$. It is thus spanned by the chain $12 - 21$ and we have

$$\begin{aligned} M_{\{1,3\}} \phi'_3 (12 - 21) &= \varrho_2[(312 + 132 + 123) - (321 + 231 + 213)] \\ &= 31 + 2.13 - 2.31 - 13 \\ &= 13 - 31 \\ &\neq 0. \end{aligned}$$

This is due to the fact that when deleting 2 in 132 and 123 (or in 321 and 231), we obtain twice the same result. More generally, it is easy to see that for $A = [\![n]\!] \setminus \{a\}$ and $B = [\![n]\!] \setminus \{b\}$ with $(a, b) \in [\![n]\!]^2$, $a \neq b$, and $\pi \in \mathfrak{S}'_A$,

$$M_B \phi'_n \pi = \phi'_B \varrho_b \pi + \pi_{b \to a},$$

where ϕ'_B is the linear operator $L(\Gamma_n) \to L(\Gamma(B))$ defined on the Dirac functions by $\pi \mapsto \mathbb{1}_{\mathfrak{S}'_B(\pi)}$ if $c(\pi) \subset B$ and 0 otherwise, and $\pi_{b \to a}$ is the word obtained when replacing b by a in π. This implies that for $x \in H_A$,

$$M_B \phi'_n x = \phi'_B \varrho_b x + \sum_{\pi \in \mathfrak{S}'_A} x(\pi) \pi_{b \to a} = \sum_{\pi \in \mathfrak{S}'_A} x(\pi) \pi_{b \to a},$$

because, since $b \in A$, $\varrho_b x = 0$ by definition of H_A. Now, it is clear that the mapping defined on the Dirac functions by $\pi \mapsto \pi_{b \to a}$ induces a bijection from $L(\Gamma(A))$ to $L(\Gamma(B))$. So if $x \neq 0$, then $M_B \phi'_n x \neq 0$. This extends to any couple of subsets $(A, B) \in \mathcal{P}([\![n]\!])^2$ such that $B \not\subset A$, and implies that we cannot take ϕ'_n as embedding operator.

3.2 Construction of W_A

The definition of our embedding operator ϕ_n requires a supplementary definition. A contiguous subword of a word $\omega = \omega_1 \dots \omega_k \in \Gamma_n$ is an expression $\omega_i \omega_{i+1} \dots \omega_{i+j}$, with $1 \leq i < i+j \leq k$. For $(A, B) \in \mathcal{P}([\![n]\!])^2$ with $A \subset B$ and $\pi \in \mathfrak{S}'_A$, we denote by $\mathfrak{S}'_B[\pi]$ the set of all the words $\sigma \in \Gamma(B)$ that contain π as a contiguous subword. For $B = [\![n]\!]$, we denote it by $\mathfrak{S}_n[\pi]$ instead of $\mathfrak{S}'_{[\![n]\!]}[\pi]$. A contiguous subword being *a fortiori* a subword, $\mathfrak{S}'_B[\pi] \subset \mathfrak{S}'_B(\pi)$.

Definition 2 (Embedding operator ϕ_n and space W_A)**.** Let ϕ_n be the linear operator $L(\Gamma_n) \to L(\mathfrak{S}_n)$ defined on Dirac functions by

$$\phi_n : \pi \mapsto \mathbb{1}_{\mathfrak{S}_n[\pi]} = \sum_{\sigma \in \mathfrak{S}_n[\pi]} \sigma,$$

and for $A \in \mathcal{P}([\![n]\!])$, let W_A be the image of H_A by ϕ_n, *i.e.*

$$W_A = \phi_n(H_A).$$

Proposition 2 (Information localization)**.** *For $A \in \mathcal{P}([\![n]\!])$, W_A satisfies conditions (7) and (8):*

$$W_A \cap \ker M_A = \{0\} \qquad \textit{and} \qquad W_A \subset \bigcap_{\substack{B \in \mathcal{P}([\![n]\!]) \\ B \not\supset A}} \ker M_B.$$

Proposition 2 is the major result of this subsection. Not only does it show that spaces W_A satisfy the good information localization properties, but it is also one of the key results to prove our multiresolution decomposition of $L(\mathfrak{S}_n)$. Its proof relies on the combinatorial properties of operator ϕ_n and requires some additional definitions.

Definition 3 (Concatenation product)**.** The concatenation product of two injective words $\pi = a_1 \dots a_r$ and $\pi' = b_1 \dots b_s$ such that $c(\pi) \cap c(\pi') = \emptyset$ is the word $\pi\pi' = a_1 \dots a_r b_1 \dots b_s$. It is extended as the bilinear operator $L(\Gamma_n) \times L(\Gamma_n) \to L(\Gamma_n)$ defined on Dirac functions by

$$(\pi, \pi') \mapsto \begin{cases} \pi\pi' & \text{if } c(\pi) \cap c(\pi') = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Starting from a word $\pi \in \Gamma_n$, the words of $\mathfrak{S}'_B[\pi]$ for $B \in \mathcal{P}([\![n]\!])$ with $c(\pi) \subset B$ are obtained by inserting the elements of $B \setminus c(\pi)$ in all possible ways, either before or after π, but not inside. Thus it is clear that

$$\mathfrak{S}'_B[\pi] = \{\omega_1 \pi \omega_2 \mid (\omega_1, \omega_2) \in \Gamma(B)^2,\ c(\omega_1) \sqcup c(\omega_2) = B \setminus c(\pi)\}$$

and $|\mathfrak{S}'_B[\pi]| = (|B| - |\pi| + 1)!$.

Example 6.

$$\mathfrak{S}_5[143] = \{25143, 52143, 21435, 51432, 14325, 14352\}.$$

The concatenation product for chains allows us to give an even simpler formula for the indicator function of the set $\mathfrak{S}'_B[\pi]$: $\mathbb{1}_{\mathfrak{S}'_B[\pi]} = \{\omega_1 \pi \omega_2 \mid (\omega_1, \omega_2) \in \Gamma(B)^2\}$. For $\omega \in \Gamma_n$, let $\mathfrak{i}_\omega$ and $\mathfrak{j}_\omega$ be the two operators on $L(\Gamma_n)$ defined on the Dirac functions by

$$\mathfrak{i}_\omega : \pi \mapsto \omega\pi \qquad \text{and} \qquad \mathfrak{j}_\omega : \pi \mapsto \pi\omega. \tag{10}$$

Operator $\mathfrak{i}_\omega$ is simply the insertion of the word ω at the beginning and $\mathfrak{j}_\omega$ at the end. It is clear that they commute and that for all $\pi \in \Gamma_n$, $\mathbb{1}_{\mathfrak{S}'_B[\pi]} = \sum_{\omega_1,\omega_2\in\Gamma(B)^2} \mathfrak{i}_{\omega_1}\mathfrak{j}_{\omega_2}\pi$. This formulation shows that the embedding operator ϕ_n is simply the sum of operators $\mathfrak{i}_{\omega_1}\mathfrak{j}_{\omega_2}$ for all $(\omega_1,\omega_2) \in (\Gamma_n)^2$:

$$\phi_n = \sum_{\omega_1,\omega_2\in\Gamma_n} \mathfrak{i}_{\omega_1}\mathfrak{j}_{\omega_2}. \tag{11}$$

Now, the proof of proposition 2 relies on this simple but crucial lemma. The technical proof can be found in the Appendix section.

Lemma 3. *For $\omega \in \Gamma_n$ and $a \in [\![n]\!] \setminus c(\omega)$,*

$$\varrho_a \mathfrak{i}_\omega = \mathfrak{i}_\omega \varrho_a \qquad \text{and} \qquad \varrho_a \mathfrak{j}_\omega = \mathfrak{j}_\omega \varrho_a.$$

Proof of proposition 2. Let $A \in \mathcal{P}([\![n]\!])$, $f \in W_A$ and $x \in H_A$ such that $f = \phi_n x$. We have

$$M_A f = M_A \phi_n x = \sum_{\pi\in\mathfrak{S}'_A} x(\pi) M_A \mathbb{1}_{\mathfrak{S}_n[\pi]} = (n-|A|+1)! \sum_{\pi\in\mathfrak{S}'_A} x(\pi)\pi = (n-|A|+1)!\; x,$$

because $\mathfrak{S}_n[\pi] \subset \mathfrak{S}_n(\pi)$ and $|\mathfrak{S}_n[\pi]| = (n-|A|+1)!$ for all $\pi \in \mathfrak{S}'_A$. Therefore if $f \in \ker M_A$, $x = 0$ and so $f = 0$. This proves that $W_A \cap \ker M_A = \{0\}$. To prove the second part, first observe that if $c(\omega) \cap A \neq \emptyset$, $\mathfrak{i}_\omega \pi = 0$ for all $\pi \in \mathfrak{S}'_A$ and thus $\mathfrak{i}_\omega x = 0$ (equivalently, $\mathfrak{j}_\omega x = 0$). Hence, using equation (11), we have

$$\phi_n x = \sum_{\omega_1,\omega_2\in\Gamma_n} \mathfrak{i}_{\omega_1}\mathfrak{j}_{\omega_2} x = \sum_{\omega_1,\omega_2\in\Gamma([\![n]\!]\setminus A)} \mathfrak{i}_{\omega_1}\mathfrak{j}_{\omega_2} x.$$

Now, let $B \in \mathcal{P}([\![n]\!])$ such that $B \not\supset A$. We want to show that $M_B f = 0$, *i.e.* that $\varrho_{[\![n]\!]\setminus B}\phi_n x = 0$. Since $B \not\supset A$, there exists $a \in A$ such that $a \notin B$, and we can write $\varrho_{[\![n]\!]\setminus B} = \varrho_{B'}\varrho_a$. Then using lemma 3,

$$\varrho_{[\![n]\!]\setminus B}\phi_n x = \varrho_{B'} \sum_{\omega_1,\omega_2\in\Gamma([\![n]\!]\setminus A)} \varrho_a \mathfrak{i}_{\omega_1}\mathfrak{j}_{\omega_2} x = \varrho_{B'} \sum_{\omega_1,\omega_2\in\Gamma([\![n]\!]\setminus A)} \mathfrak{i}_{\omega_1}\mathfrak{j}_{\omega_2}\varrho_a x = 0,$$

because $\varrho_a x = 0$ by definition of H_A. □

3.3 The decomposition of $L(\mathfrak{S}_n)$

Now that we have constructed the subspaces of $L(\mathfrak{S}_n)$ that localize the information specific to each marginal, we show that they constitute a decomposition of the space $L(\mathfrak{S}_n)$. Recall that V^0 is the subspace of $L(\mathfrak{S}_n)$ of constant functions. So defining $L_0(\mathfrak{S}_n) = \{f \in L(\mathfrak{S}_n) \mid \sum_{\sigma\in\mathfrak{S}_n} f(\sigma) = 0\}$, we have

$$L(\mathfrak{S}_n) = V^0 \oplus L_0(\mathfrak{S}_n). \tag{12}$$

Proposition 3. *The spaces $(W_A)_{A\in\mathcal{P}([\![n]\!])}$ are in direct sum in $L_0(\mathfrak{S}_n)$.*

Proof. First, observe that for $A \in \mathcal{P}([\![n]\!])$ and $x \in H_A$,

$$\sum_{\sigma\in\mathfrak{S}_n} (\phi_n x)(\sigma) = \sum_{\sigma\in\mathfrak{S}_n} \sum_{\pi\in\mathfrak{S}'_A} x(\pi)\mathbb{1}_{\mathfrak{S}_n[\pi]}(\sigma) = (n-|A|+1)! \sum_{\pi\in\mathfrak{S}'_A} x(\pi) = 0,$$

because as $x \in H_A$, $0 = \varrho_A x = \left[\sum_{\pi \in \mathfrak{S}'_A} x(\pi)\right] \overline{0}$. Hence, $W_A \subset L_0(\mathfrak{S}_n)$. To prove that the spaces W_A are in direct sum, let $(f_A)_{A \in \mathcal{P}([\![n]\!])}$ be a family of functions with $f_A \in W_A$ for each $A \in \mathcal{P}([\![n]\!])$, such that

$$\sum_{A \in \mathcal{P}([\![n]\!])} f_A = 0. \tag{13}$$

We need to show that $f_A = 0$ for all $A \in \mathcal{P}([\![n]\!])$. We proceed by induction on the cardinality of A. Let $A \in \binom{[\![n]\!]}{2}$. For all $B \in \mathcal{P}([\![n]\!])$ different from A, we have $A \not\supset B$. Thus, using the second part of proposition 2, $M_A f_B = 0$ for all $B \in \mathcal{P}([\![n]\!]) \setminus \{A\}$. Applying M_A in equation (13) then gives $M_A f_A = 0$. This means that $f_A \in W_A \cap \ker M_A$ and so that $f_A = 0$, using the first part of proposition 2. Now assume that $f_A = 0$ for all $A \in \mathcal{P}([\![n]\!])$ such that $|A| \leq k-1$, with $k \in \{3, \dots, n\}$. Equation (13) then becomes

$$\sum_{|A| \geq k} f_A = 0. \tag{14}$$

Let $A \in \binom{[\![n]\!]}{k}$. For all $B \subset [\![n]\!]$ such that $|B| \geq k$ and different from A, we have $A \not\supset B$. Thus, using again proposition 2, $M_A f_B = 0$, and applying this to equation (14) gives $M_A f_A$. We conclude using proposition 2 one more time. □

The second step in the proof of our decomposition is a dimensional argument. Notice that for $A \in \binom{[\![n]\!]}{k}$ with $k \in \{2, \dots, n\}$, H_A is isomorphic to the space

$$H_k = \{x \in L(\Gamma(\{1, \dots, k\})) \mid \varrho_i x = 0 \text{ for all } i \in \{1, \dots, k\}\}.$$

Now, it happens that this space is actually closely related to another well-studied space in the algebraic topology literature, namely the top homology space of the complex of injective words (see [8], [3], [36], [12]). The link is made in [35] (the space H_k is denoted by $\ker \pi_{[\![k]\!]}$), and leads in particular to the following result (see proposition 6.8 and corollary 6.15).

Theorem 1 (Dimension of H_k)**.** *For $k \in \{2, \dots, n\}$,*

$$\dim H_k = d_k,$$

where d_k is the number of fixed-point free permutations (also called derangements) on the set $\{1, \dots, k\}$.

As simple as it may seem, this result is far from being trivial. Its proof relies on the topological nature of the partial order of subword inclusion on the complex of injective words and the use of the Hopf trace formula for virtual characters. It is a cornerstone in the proof of our multiresolution decomposition.

Theorem 2 (Multiresolution decomposition)**.** *The following decomposition of $L(\mathfrak{S}_n)$ holds:*

$$L(\mathfrak{S}_n) = V^0 \oplus \bigoplus_{A \in \mathcal{P}_2([\![n]\!])} W_A.$$

In addition, $\dim W_A = d_{|A|}$ *and* $\phi_n(H_A) = W_A$ *for all* $A \in \mathcal{P}([\![n]\!])$.

Proof. For $A \in \mathcal{P}([\![n]\!])$ and $x \in H_A$, $\phi_n x = \sum_{\pi \in \mathfrak{S}'_A} x(\pi) \mathbb{1}_{\mathfrak{S}_n[\pi]}$ by definition. Since for $(\pi, \pi') \in (\mathfrak{S}'_A)^2$ such that $\pi \neq \pi'$, the sets $\mathfrak{S}_n[\pi]$ and $\mathfrak{S}_n[\pi']$ are disjoint, it is clear that $\phi_n x = 0 \Rightarrow x(\pi) = 0$ for all $\pi \in \mathfrak{S}'_A$, *i.e.* $x = 0$. This proves that the restriction of ϕ_n to H_A is injective, and thus that $\dim W_A \geq \dim H_A$, *i.e.* $\dim W_A \geq d_{|A|}$, using theorem 1. Now, using proposition 3 and equation (12), we obtain

$$\dim \left[V^0 \oplus \bigoplus_{A \in \mathcal{P}_2([\![n]\!])} W_A \right] \geq 1 + \sum_{k=2}^{n} \binom{n}{k} d_k = \sum_{k=0}^{n} \binom{n}{k} d_{n-k} = n!,$$

where the last equality results from the observation that the number of permutations with k fixed points is equal to $\binom{n}{k} d_{n-k}$. Since $\dim L(\mathfrak{S}_n) = n!$, this concludes both the proof of the decomposition of $L(\mathfrak{S}_n)$ and the dimension of W_A, and the fact that $\phi_n(H_A) = W_A$ follows. □

This decomposition appears implicitly in [35], in the combination of theorem 6.20 and formula (22). It is however defined modulo isomorphism, and not easily usable for applications. Our explicit construction permits a practical use of this decomposition. In particular, it allows to localize the information related to any observation design $\mathcal{A} \subset \mathcal{P}([\![n]\!])$, as declared in the introduction.

Corollary 1. *For any subset $A \in \mathcal{P}([\![n]\!])$,*

$$L(\mathfrak{S}_n) = \ker M_A \oplus \left[V^0 \oplus \bigoplus_{B \in \mathcal{P}(A)} W_B \right],$$

and for any observation design $\mathcal{A} \subset \mathcal{P}([\![n]\!])$,

$$L(\mathfrak{S}_n) = \ker M_{\mathcal{A}} \oplus \left[V^0 \oplus \bigoplus_{B \in \bigcup_{A \in \mathcal{A}} \mathcal{P}(A)} W_B \right].$$

Proof. Let $A \in \mathcal{P}([\![n]\!])$. By theorem 2, we have

$$\ker M_A = \left(\ker M_A \cap V^0 \right) \oplus \bigoplus_{B \in \mathcal{P}([\![n]\!])} \left(\ker M_A \cap W_B \right).$$

It is clear that $\ker M_A \cap V^0 = \{0\}$ (M_A maps constant functions on $\mathfrak{S}_n$ to constant functions on $\mathfrak{S}'_A$). Moreover, for $B \in \mathcal{P}(A)$, $\ker M_A \cap W_B \subset \ker M_B \cap W_B = \{0\}$, using proposition 1 and the first part of proposition 2. At last, for all $B \in \mathcal{P}([\![n]\!]) \setminus \mathcal{P}(A)$, $A \not\supset B$, and thus $W_B \subset \ker M_A$, using the second part of proposition 2. This means that $\ker M_A = \bigoplus_{B \in \mathcal{P}([\![n]\!]) \setminus \mathcal{P}(A)} W_B$, and the first part of corollary 1 follows. The second part results from the calculation

$$\ker M_{\mathcal{A}} = \bigcap_{A \in \mathcal{A}} \ker M_A = \bigcap_{A \in \mathcal{A}} \bigoplus_{B \in \mathcal{P}([\![n]\!]) \setminus \mathcal{P}(A)} W_B = \bigoplus_{B \in \mathcal{P}([\![n]\!]) \setminus \bigcup_{A \in \mathcal{A}} \mathcal{P}(A)} W_B.$$

□

Example 7. Let's consider an example with $n = 4$. The multiresolution decomposition of $L(\mathfrak{S}_4)$ is given by the following representation.

$$
\begin{gathered}
W_{\{1,2,3,4\}} \\
W_{\{1,2,3\}} \oplus W_{\{1,2,4\}} \oplus W_{\{1,3,4\}} \oplus W_{\{2,3,4\}} \\
W_{\{1,2\}} \oplus W_{\{1,3\}} \oplus W_{\{1,4\}} \oplus W_{\{2,3\}} \oplus W_{\{2,4\}} \oplus W_{\{3,4\}} \\
V^0
\end{gathered}
$$

The spaces in bold contain the information related to the observation of marginals on $\{1,2,3\}$ in this representation,

$$
\begin{gathered}
W_{\{1,2,3,4\}} \\
\mathbf{W_{\{1,2,3\}}} \oplus W_{\{1,2,4\}} \oplus W_{\{1,3,4\}} \oplus W_{\{2,3,4\}} \\
\mathbf{W_{\{1,2\}}} \oplus \mathbf{W_{\{1,3\}}} \oplus W_{\{1,4\}} \oplus \mathbf{W_{\{2,3\}}} \oplus W_{\{2,4\}} \oplus W_{\{3,4\}} \\
\mathbf{V^0}
\end{gathered}
$$

to the observation of marginals on $\{1,3,4\}$ in this one,

$$
\begin{gathered}
W_{\{1,2,3,4\}} \\
W_{\{1,2,3\}} \oplus W_{\{1,2,4\}} \oplus \mathbf{W_{\{1,3,4\}}} \oplus W_{\{2,3,4\}} \\
W_{\{1,2\}} \oplus \mathbf{W_{\{1,3\}}} \oplus \mathbf{W_{\{1,4\}}} \oplus W_{\{2,3\}} \oplus W_{\{2,4\}} \oplus \mathbf{W_{\{3,4\}}} \\
\mathbf{V^0}
\end{gathered}
$$

and to the observation of marginals of the observation design $\{\{1,2,3\},\{1,3,4\}\}$ in this final representation.

$$
\begin{gathered}
W_{\{1,2,3,4\}} \\
\mathbf{W_{\{1,2,3\}}} \oplus W_{\{1,2,4\}} \oplus \mathbf{W_{\{1,3,4\}}} \oplus W_{\{2,3,4\}} \\
\mathbf{W_{\{1,2\}}} \oplus \mathbf{W_{\{1,3\}}} \oplus \mathbf{W_{\{1,4\}}} \oplus \mathbf{W_{\{2,3\}}} \oplus W_{\{2,4\}} \oplus \mathbf{W_{\{3,4\}}} \\
\mathbf{V^0}
\end{gathered}
$$

From a practical point of view, if we observe $(f_A)_{A\in\mathcal{A}} \in \mathbb{M}_{\mathcal{A}}$ then by corollary 1, there exists a unique $f \in V^0 \oplus \bigoplus_{B\in\bigcup_{A\in\mathcal{A}}\mathcal{P}(A)} W_B$ such that $M_{\mathcal{A}} f = (f_A)_{A\in\mathcal{A}}$. Furthermore, if

$$
f = \tilde{f}_0 + \sum_{B\in\bigcup_{A\in\mathcal{A}}\mathcal{P}(A)} \tilde{f}_B
$$

is the decomposition of f corresponding to $\bigoplus_{B\in\bigcup_{A\in\mathcal{A}}\mathcal{P}(A)} W_B$, we obtain the wanted relation (6): for any $A \in \mathcal{A}$,

$$
f_A = M_A \left[\tilde{f}_0 + \sum_{B\in\mathcal{P}(A)} \tilde{f}_B \right].
$$

3.4 Multiresolution analysis

Until now, we have only used the expression "multiresolution decomposition", not "multiresolution analysis". The latter has indeed a specific mathematical definitions, first formalized in [30] and [27] (it is called "multiresolution approximation" in the latter) for the space $L^2(\mathbb{R})$. A multiresolution analysis of $L^2(\mathbb{R})$ is a sequence $(\tilde{V}^j)_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ such that:

1. $\tilde{V}^j \subset \tilde{V}^{j+1}$ for all $j \in \mathbb{Z}$
2. $\bigcup_{j\in\mathbb{Z}} \tilde{V}^j = L^2(\mathbb{R})$ and $\bigcap_{j\in\mathbb{Z}} \tilde{V}^j = \{0\}$
3. $f(x) \in \tilde{V}^j \Leftrightarrow f(2x) \in \tilde{V}^{j+1}$ for all $j \in \mathbb{Z}$
4. $f(x) \in \tilde{V}^j \Leftrightarrow f(x - 2^{-j}k) \in \tilde{V}^j$ for all $k \in \mathbb{Z}$
5. There exists $g \in \tilde{V}^0$ such that $(g(x-k))_{k\in\mathbb{Z}}$ is a Riesz basis of $\tilde{V}^0$.

In order to define an analogous definition for $L(\mathfrak{S}_n)$, we get back to the general principles behind it. The idea is that the index j represents a scale, and each space $\tilde{V}^j$ contains the information of all scales lower than j, thus $\tilde{V}^j \subset \tilde{V}^{j+1}$. In finite dimension, the number of scales is necessarily finite, and we request that the space of largest scale be equal to the full space (we can request that the space of lower scale be $\{0\}$ but it is useless). The principle of multiresolution analysis is not only to have a nested sequence of subspaces corresponding to different scales, but also to define the operators that leave a space $\tilde{V}^j$ invariant and the ones that send from a space $\tilde{V}^j$ to $\tilde{V}^{j+1}$ and vice versa. In the case of $L^2(\mathbb{R})$, these operators are respectively the scaled translation $f(x) \mapsto f(x - 2^{-j}k)$ and the dilation $f(x) \mapsto f(2x)$, defined in conditions 3. and 4. If we see a function f as an image, the dilation corresponds to a zoom, and a scaled translation corresponds to a displacement.

To define a multiresolution analysis in our case, we first need a notion of scale. In our construction, the natural notion of scale for the spaces W_A's appears clearly on the precedent representations of the multiresolution decomposition of $L(\mathfrak{S}_n)$: the cardinality of the indexing subsets. We say that the marginal p_A of a probability distribution p on $\mathfrak{S}_n$ on a subset $A \in \mathcal{P}(\llbracket n \rrbracket)$ is of scale k if $|A| = k$. This means that p_A is a probability distribution over rankings involving k items. The information contained in p_A can be decomposed in components of scales $\leq k$, and the projection of p on W_A contains the information of scale k. For $k \in \{2, \dots, n\}$, we define the space W^k that contains all the information of scale k by

$$W^k = \bigoplus_{|A|=k} W_A \tag{15}$$

and the space V^k that contains all the information of scales $\leq k$ by

$$V^k = V^0 \oplus \bigoplus_{j=2}^{k} \bigoplus_{|A|=j} W_A. \tag{16}$$

We thus have

$$V^0 \subset V^2 \subset V^3 \subset \cdots \subset V^n = L(\mathfrak{S}_n) \qquad \text{and} \qquad L(\mathfrak{S}_n) = V^0 \oplus \bigoplus_{k=2}^{n} W^k. \tag{17}$$

The space W^k represent the information gained at scale k, and for that we can call it a detail space by analogy with multiresolution analysis on $L^2(\mathbb{R})$. In the present case however, the decomposition of $L(\mathfrak{S}_n)$ into detail spaces is not orthogonal, as we shall see in section 4.

The second step in the definition of a multiresolution analysis is the construction of operators of "zoom" and "displacement". Translations on $\mathbb{R}$ are given by the additive action of $\mathbb{R}$ on itself, *i.e.* are of the form $x \mapsto x + a$, with $a \in \mathbb{R}$. The associated translations on $L^2(\mathbb{R})$ are then of the

form $f \mapsto f(.-a)$, so that the indicator function of a singleton $\{x\}$ is sent to the indicator function of the singleton $\{x+a\}$. In the case of injective words, we consider the canonical action of $\mathfrak{S}_n$ on Γ_n, defined by $\pi \mapsto \sigma_0(\pi)$, where for $\sigma_0 \in \mathfrak{S}_n$ and $\pi = \pi_1 \dots \pi_k \in \Gamma_n$, $\sigma_0(\pi)$ is the injective word $\sigma_0(\pi_1)\dots\sigma_0(\pi_k)$. We then define the associated translations on $L(\Gamma_n)$ as the linear operators T_{σ_0} defined on Dirac functions by

$$T_{\sigma_0}\pi = \sigma_0(\pi), \tag{18}$$

for $\sigma_0 \in \mathfrak{S}_n$. We could still denote the translation operator by σ_0 but we choose the notation T_{σ_0} for clarity's sake. It is easy to see that the orbits of the action are the Γ^k, for $k \in \{0,\dots,n\}$. Translation operators thus stabilize each space $L(\Gamma^k)$, and in particular $L(\Gamma^n) = L(\mathfrak{S}_n)$. We still denote by T_{σ_0} the induced operator. By construction, the operator (and its induced operators) T_{σ_0} is invertible with inverse $T_{\sigma_0}^{-1} = T_{\sigma_0^{-1}}$, for any $\sigma_0 \in \mathfrak{S}_n$.

Remark 2. If $\pi = \pi_1 \dots \pi_n \in \Gamma^n$ is seen as a permutation, then $\pi_i = \pi^{-1}(i)$ for $i \in [\![n]\!]$ and $\sigma_0(\pi)$ is the injective word associated to the permutation $\pi\sigma_0^{-1}$. Translation T_{σ_0} on $L(\Gamma^n) = L(\mathfrak{S}_n)$ can thus also be defined by $T_{\sigma_0}f(\sigma) = f(\sigma\sigma_0)$. The mapping $\sigma_0 \mapsto T_{\sigma_0}$ is called the right regular representation in group representation theory.

Lemma 4. *Let $\sigma_0 \in \mathfrak{S}_n$, $\omega \in \Gamma_n$ and $a \in [\![n]\!]$.*

1. $T_{\sigma_0}\varrho_a = \varrho_{\sigma_0(a)}T_{\sigma_0}$.
2. $T_{\sigma_0}\mathfrak{i}_\omega = \mathfrak{i}_{\sigma_0(\omega)}T_{\sigma_0}$ *and* $T_{\sigma_0}\mathfrak{j}_\omega = \mathfrak{j}_{\sigma_0(\omega)}T_{\sigma_0}$.
3. $T_{\sigma_0}\phi_n = \phi_n T_{\sigma_0}$, *i.e.* $T_{\sigma_0}\mathbb{1}_{\mathfrak{S}_n[\pi]} = \mathbb{1}_{\mathfrak{S}_n[\sigma_0(\pi)]}$ *for all* $\pi \in \Gamma_n$.

Proof. Properties 1. and *2.* are trivially verified. To prove *3.*, observe that $\omega \mapsto \sigma_0(\omega)$ being a group action, it is bijective, and thus using equation (11) and *2.*, we obtain

$$T_{\sigma_0}\phi_n = \sum_{\omega_1,\omega_2\in\Gamma_n} T_{\sigma_0}\mathfrak{i}_{\omega_1}\mathfrak{j}_{\omega_2} = \sum_{\omega_1,\omega_2\in\Gamma_n} \mathfrak{i}_{\sigma_0(\omega_1)}\mathfrak{j}_{\sigma_0(\omega_2)}T_{\sigma_0} = \left[\sum_{\omega_1',\omega_2'\in\Gamma_n} \mathfrak{i}_{\omega_1'}\mathfrak{j}_{\omega_2'}\right] T_{\sigma_0} = \phi_n T_{\sigma_0}.$$

□

The following proposition shows that translation operators T_{σ_0} can be seen as "displacement" operators adapted to our multiresolution decomposition.

Proposition 4 (Displacement operator)**.** *Let $k \in \{2,\dots,[\![n]\!]\}$, $(A,B) \in \binom{[\![n]\!]}{k}^2$ and $\sigma_0 \in \mathfrak{S}_n$ such that $\sigma_0(A) = B$. Then*

$$T_{\sigma_0}(W_A) = W_B.$$

Proof. Since T_{σ_0} is invertible and $\dim W_A = \dim W_B = d_k$ by theorem 2, we only need to prove that $T_{\sigma_0}(W_A) \subset W_B$. Let $x \in H_A$. Property *3.* in lemma 4 gives $T_{\sigma_0}\phi_n x = \phi_n T_{\sigma_0}x$. Thus we just have to show that $T_{\sigma_0}x \in H_B$. Since σ_0 is a permutation such that $\sigma_0(A) = B$, it is clear that $\{\sigma_0(\pi) \mid \pi \in \mathfrak{S}'_A\} = \mathfrak{S}'_B$, and $T_{\sigma_0}x \in L(\mathfrak{S}'_B)$. Now, using property *1.* in lemma 4, we have for any $b \in B$, $\varrho_b T_{\sigma_0}x = T_{\sigma_0}\varrho_{\sigma_0^{-1}(b)}x = 0$ because $\sigma_0^{-1}(b) \in A$. □

Looking at definitions (15) and (16), proposition 4 immediately gives the following result.

Proposition 5 (Translation invariance)**.** *For $k \in \{2,\dots,n\}$, the spaces W^k and V^k are invariant under all the translations T_{σ_0}, for $\sigma_0 \in \mathfrak{S}_n$.*

Observe that the space V^j is invariant under all translations T_{σ_0} whereas in the case of the multiresolution analysis on $L^2(\mathbb{R})$, the space $\tilde{V}^j$ is only invariant under scaled translations $f \mapsto f(. - 2^{-j}k)$. The latter property means that the size of translations is limited by the resolution level. The same interpretation is actually also true in our context: though V^j is invariant under all translations T_{σ_0}, they only involve the action of $\mathfrak{S}_n$ on the sets Γ^i for $i \leq j$. The "size" of translations on V^j is thus inherently limited by the resolution level.

While the construction of our displacement operator is based on the same algebraic objects as for $L^2(\mathbb{R})$, namely translations associated to a group action, it is not possible to base the construction of a zooming operator on dilation. This is the bottleneck of any construction of a multiresolution analysis on a discrete space such as $\mathfrak{S}_n$, as observed in [23]. Hence, there is no simple way to define an operator that allows to change scales, such as $f(x) \mapsto f(2x)$. We can however construct a family of "dezooming" operators Φ_k that each project onto the corresponding space V^k. For $k \in \{2, \dots, n\}$, we denote by M_k the operator associated to all the marginals of scale k, *i.e.* $M_k := M_{\binom{[\![n]\!]}{k}}$,

$$\begin{aligned} M_k : L(\mathfrak{S}_n) &\to \bigoplus_{|A|=k} L(\mathfrak{S}'_A) \\ f &\mapsto (f_A)_{|A|=k}. \end{aligned}$$

Using corollary 1 for $\mathcal{A} = \binom{[\![n]\!]}{k}$, we have

$$L(\mathfrak{S}_n) = \ker M_k \oplus \left[V^0 \oplus \bigoplus_{B \in \bigcup_{|A|=k} \mathcal{P}(A)} W_B \right] = \ker M_k \oplus V^k.$$

Therefore, for any $F \in M_k(L(\mathfrak{S}_n))$, there exists a unique $f \in V^k$ such that $M_k f = F$. We denote by M_k^+ the operator from $M_k(L(\mathfrak{S}_n))$ to V^k that sends F to f. This is a pseudoinverse of M_k, but not the Moore-Penrose pseudoinverse because V^k is not the orthogonal supplementary of $\ker M_k$.

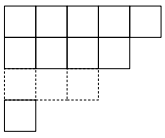
Definition 4 (Dezooming operator)**.** Let $\Phi_0 : f \mapsto (\langle f, \mathbb{1}_{\mathfrak{S}_n} \rangle / n!)\mathbb{1}_{\mathfrak{S}_n}$ be the orthogonal projection on V^0 and for $k \in \{2, \dots, n\}$,

$$\Phi_k = M_k^+ M_k.$$

3.5 Decomposition of the space W^k into irreducible components

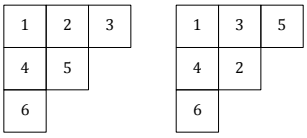
By proposition 5, the space W^k with $k \in \{2, \dots, n\}$ is invariant under all the translations T_{σ_0} for all $\sigma_0 \in \mathfrak{S}_n$. In other words, it is a representation of the symmetric group $\mathfrak{S}_n$. It can thus be decomposed as a sum of irreducible representations S^λ. The multiplicity of each irreducible is nonetheless not obvious to compute. This is one of the major results established in [35]. Its statement requires some definitions.

A Young diagram (or a Ferrer's diagram) of size n is a collection of boxes of the form



where if λ_i denotes the number of boxes in row i, then $\lambda = (\lambda_1, \ldots, \lambda_r)$, called the shape of the Young diagram, must be a partition of n. The total number of boxes of a Young diagram is therefore equal to n, and each row contains at most as many boxes as the row above it. A Young tableau is a Young diagram filled with all the integers $1, \ldots, n$, one in each boxes. The shape of a Young tableau Q, denoted by $\text{shape}(Q)$, is the shape of the associated Young Diagram, it is thus a partition of n. There are clearly $n!$ Young tableaux of a given shape $\lambda \vdash n$. A Young tableau is said to be *standard* if the numbers increase along the rows and down the columns.

Example 8. In the following figure, the first tableau is standard whereas the second is not.



Notice that a standard Young tableau always have 1 in its top-left box, and that the box that contains n is necessarily at the end of a row and a column. We denote by SYT_n the set of all standard Young tableaux of size n and by $\text{SYT}_n(\lambda) = \{Q \in \text{SYT}_n \mid \text{shape}(Q) = \lambda\}$ the set of standard Young tableaux of shape λ, for $\lambda \vdash n$. By construction, $\text{SYT}_n = \bigsqcup_{\lambda \vdash n} \text{SYT}_n(\lambda)$. A classic result in the representation theory of the symmetric group is that for all $\lambda \vdash n$, the dimension d_λ of S^λ, which is also its multiplicity in the decomposition of $L(\mathfrak{S}_n)$, is actually equal to the number of standard Young tableaux of shape λ. Thus the decomposition of $L(\mathfrak{S}_n)$ into irreducible representations is given by:

$$L(\mathfrak{S}_n) = \bigoplus_{Q \in \text{SYT}_n} S^{\text{shape}(Q)}.$$

Figure 7 represents all the standard Young tableaux of size $n = 4$, gathered by shape.

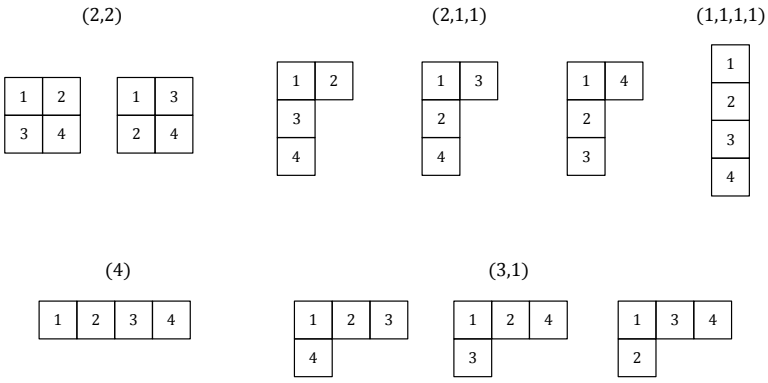
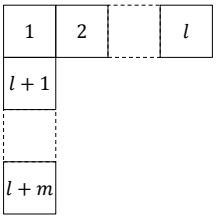


Figure 7: Standard Young tableaux of size $n = 4$

By construction, $L(\mathfrak{S}_n) = V^0 \oplus \bigoplus_{k=2}^{n} W^k$, where V^0 is isomorphic to the Specht module $S^{(n)} = S^{\text{shape}(Q_0)}$, Q_0 being the unique standard Young tableau of shape (n). So for each $k \in$

$\{2, \dots, n\}$ the decomposition of the space W^k must involve a certain subset T_k of SYT_n, such that $\mathrm{SYT}_n = \{Q_0\} \sqcup \bigsqcup_{k=2}^{n} T_k$. The construction of these subsets is done in [35]. We reproduce it here. Let Q be a standard Young tableau. Then it contains a unique maximal subtableau of the form



with $1 \leq l \leq n$ and $0 \leq m \leq n - l$. Define

$$\mathrm{eig}(Q) = \begin{cases} l & \text{if } m \text{ is even,} \\ l-1 & \text{if } m \text{ is odd.} \end{cases} \tag{19}$$

(This definition is given in [35], in the proof of Proposition 6.23).

Theorem 3 (Decomposition of W^k into irreducible representations)**.** *For $k \in \{2, \dots, n\}$, the following decomposition holds*

$$W^k \cong \bigoplus_{\substack{Q \in SYT_n \\ eig(Q) = n-k}} S^{\,shape(Q)}.$$

Proof. For $k \in \{2, \dots, n\}$,

$$W^k = \bigoplus_{|A|=k} W_A = \bigoplus_{|A|=k} \phi_n(H_A) = \phi_n\left(\bigoplus_{|A|=k} H_A\right) \cong \bigoplus_{|A|=k} H_A,$$

where the two last linear spaces are isomorphic because ϕ_n is an isomorphism between H_A and W_A for any $A \in \mathcal{P}([\![n]\!])$ by theorem 2. Furthermore, point 3. of lemma 4 shows that W^k and $\bigoplus_{|A|=k} H_A$ are isomorphic as representations of $\mathfrak{S}_n$. In the notations of [35], $H_A = \ker \pi_A$, so by their theorem 6.20, $W^k \cong F_{n,n-k}$ as representations of $\mathfrak{S}_n$. Theorem 3 is then just a reformulation in the present setting of theorem 6.26 from [35]. □

For $k \in \{2, \dots, n\}$, the subset T_k of standard Young tableaux involved in the decomposition of W^k is thus defined by $T_k = \{Q \in \mathrm{SYT}_n \mid \mathrm{eig}(Q) = n - k\}$. Figure 8 represents the different subsets T_k with the associated decompositions for $n = 4$.

Remark 3. For $\lambda = (\lambda_1, \dots, \lambda_r) \vdash n$, the usual ranking interpretation of the Specht module S^λ is that it localizes information at "scale" $n - \lambda_1$, in the sense that it localizes the absolute rank information of $n - \lambda_1$ items. The Specht module $S^{(n-1,1)}$ localizes absolute rank information about 1 item, $S^{(n-2,2)}$ and $S^{(n-2,1,1)}$ both localize absolute rank information about 2 items, and so on. It is interesting to notice that this interpretation does not hold when dealing with relative rank information. The space W^k can indeed be seen as localizing the relative rank information at scale k *i.e.* the relative rank information related to incomplete rankings involving k items. However, theorem 3 shows that the decomposition of W^k can involve Specht modules S^λ with $n - \lambda_1 \neq k$. Figure 8 shows for example that for $n = 4$, W^3 involves absolute rank information of scale 1 and 2.

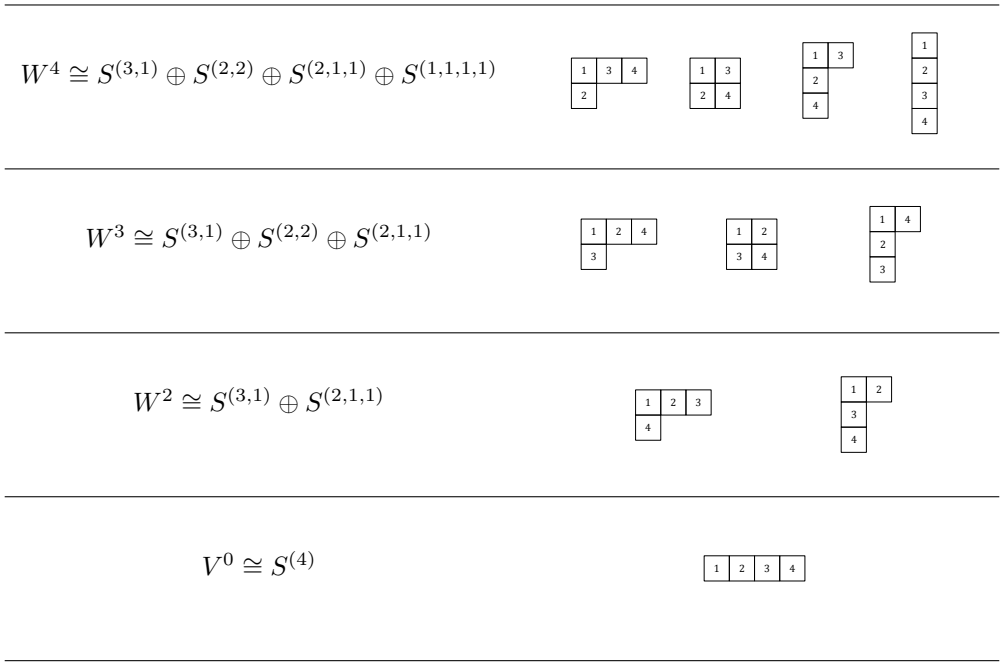


Figure 8: Spaces W^k and their decompositions into irreducibles, for $n = 4$

4 The wavelet basis

We now construct an explicit basis Ψ adapted to the multiresolution decomposition of $L(\mathfrak{S}_n)$, in the sense that $\Psi = \{\psi_0\} \cup \bigcup_{A \in \mathcal{P}(\llbracket n \rrbracket)} \Psi_A$ where Ψ_A is a basis of W_A for all $A \in \mathcal{P}(\llbracket n \rrbracket)$, and establish its main properties.

4.1 Generative algorithm

The basis is defined by an algorithm adapted from [33], which requires some definitions about cycles and permutations. A cycle on $\llbracket n \rrbracket$ is a permutation $\gamma \in \mathfrak{S}_n$ for which there exist m distinct elements $a_1, \ldots, a_m \in \llbracket n \rrbracket$, with $m \geq 2$, such that $\gamma(a_i) = a_{i+1}$ for $i = 1, \ldots, m-1$, $\gamma(a_m) = a_1$, and $\gamma(a') = a'$ for all $a' \in \llbracket n \rrbracket \setminus \{a_1, \ldots, a_m\}$. The cycle γ is then denoted by $(a_1 \ldots a_m)$, its support is the set $\{a_1, \ldots, a_m\}$ and its length is $l(\gamma) = m$. For $A \in \mathcal{P}(\llbracket n \rrbracket)$, we denote by $\mathrm{Cycle}(A)$ the set of all cycles with support A. It is well known that a permutation $\tau \in \mathfrak{S}_n$ admits a unique decomposition as a product of cycles with distinct supports $\tau = \gamma_1 \ldots \gamma_r$ (fixed-points are not represented). This decomposition can though be written in several ways, depending on the order of the cycles and the first element of each cycle.

Definition 5 (Standard cycle form)**.** A permutation is written in standard cycle form if it is written as a product of disjoint cycles so that the minimum element of a cycle appears at the leftmost letter in that cycle, and the cycles are arranged from left to right in increasing values of minimum letters.

Example 9. The permutation (134)(25) is written in standard cycle form, while the alternative representations (413)(25) or (25)(134) are not.

For a permutation $\tau \in \mathfrak{S}_n$, we denote by $\mathrm{cyc}(\tau)$ the number of its cycles, define its support by $\mathrm{supp}(\tau) = \{i \in [\![n]\!] \mid \tau(i) \neq i\}$ and its length by $l(\tau) = |\mathrm{supp}(\tau)|$. These definitions extend the definitions of the support and the length for a cycle, and if $\gamma_1 \dots \gamma_{\mathrm{cyc}(\tau)}$ is the cycle decomposition of τ, $l(\tau) = l_1 + \dots + l_{\mathrm{cyc}(\tau)}$. For $A \in \mathcal{P}([\![n]\!])$, we define

$$\mathcal{D}_A = \{\sigma \in \mathfrak{S}_n \mid \mathrm{supp}(\sigma) = A\},$$

and we set by convention $\mathcal{D}_\emptyset = \{id\}$, where $id \in \mathfrak{S}_n$ is the identity permutation on $[\![n]\!]$. By definition, a permutation $\sigma \in \mathcal{D}_A$ induces a fixed-point free permutation, also called a derangement, on A. The set $\mathcal{D}_A$ is thus the natural embedding of the set of derangements on A in $\mathfrak{S}_n$. The algorithm of [33] computes a basis for the top homology space of the complex of injective words over the field $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$ of two elements. It uses the operation on $\mathbb{F}_2$-valued chains "$x \diamond y = xy + yx$". In the present setting, we use the following definition.

Definition 6 (Diamond operator)**.** For $x, y \in L(\Gamma_n)$, we define

$$x \diamond y = xy - yx.$$

The algorithm of [33] takes a derangement on $\{1, \dots, k\}$ as input and outputs an element of the top homology space of the complex of injective words. It happens that the same algorithm with the diamond operator of definition 6 maps a derangement on $\{1, \dots, k\}$ to an element of H_k. Moreover, the algorithm is naturally extended to take a permutation $\tau \in \mathcal{D}_A$ as input and output an element x_τ of the space H_A, for any $A \in \mathcal{P}([\![n]\!])$.

Algorithm 1. Let $A \in \mathcal{P}([\![n]\!])$. The input is a permutation $\tau \in \mathcal{D}_A$ written in standard cycle form, and the output is a chain $x_\tau \in H_A$.

Step 1.	Between each consecutive pair of letters in each cycle of τ, insert the symbol $\star$.
Step 2.	If there are no $\star$ symbols in the string, then HALT. Otherwise, determine which symbol $\star$ has the largest right-hand neighbor.
Step 2.	Suppose that the symbol located in Step 2 is between quantities Q and R; that is, it appears as $Q \star R$. Then replace $Q \star R$ by $(Q \diamond R)$.
Step 4.	GOTO Step 2.

Example 10. Let $A = \{1, 2, 3, 4, 5\}$ and $\tau = (134)(25)$. Algorithm 1 gives the following sequence of steps.

$$\begin{gathered}(1 \star 3 \star 4)(2 \star 5)\\ (1 \star 3 \star 4)(2 \diamond 5)\\ (1 \star (3 \diamond 4))(2 \diamond 5)\\ (1 \diamond (3 \diamond 4))(2 \diamond 5)\end{gathered}$$

Expanding the concatenation and the $\diamond$ operations, we obtain:

$$\begin{aligned} x_{(134)(25)} &= (1 \diamond (3 \diamond 4))(2 \diamond 5)\\ &= (1 \diamond (34 - 43))(25 - 52)\\ &= (134 - 143 - 341 + 431)(25 - 52)\\ &= 13425 - 13452 - 14325 + 14352 - 34125 + 34152 + 43125 - 43152. \end{aligned}$$

4.2 The basis of $L(\mathfrak{S}_n)$

We now construct the wavelet basis of $L(\mathfrak{S}_n)$. We first show that the outputs of algorithm 1 belong to the claimed space.

Proposition 6. *Let $A \in \mathcal{P}([\![n]\!])$. For all $\tau \in \mathcal{D}_A$, $x_\tau \in H_A$.*

As in [33], the proof relies on the simple following lemma, of which proof is straightforward and is thus omitted.

Lemma 5. *Let $x, y \in L(\Gamma_n)$ with $c(x) \cap c(y) = \emptyset$, and $a \in c(x)$. Then*

$$\varrho_a(xy) = \varrho_a(x)y \qquad \text{and} \qquad \varrho_a(x \diamond y) = \varrho_a(x) \diamond y.$$

Proof of proposition 6. Let $A \in \mathcal{P}([\![n]\!])$ and $\tau \in \mathcal{D}_A$. We need to show that for all $a \in A$, $\varrho_a x_\tau = 0$. Let $a \in A$ and $\tau = \gamma_1 \dots \gamma_r$ be the standard cycle form of τ. By definition of $\mathcal{D}_A$, $\{\mathrm{supp}(\gamma_1), \dots, \mathrm{supp}(\gamma_r)\}$ is a partition of A. Let γ_i be the cycle which support contains a. By definition of the algorithm, $x_\tau = x_{\gamma_1} \dots x_{\gamma_r}$, and using lemma 5, we have $\varrho_a(x_\tau) = x_{\gamma_1} \dots \varrho_a(x_{\gamma_i}) \dots x_{\gamma_r}$. Since γ_i is a cycle, its support contains at least two elements, and thus x_{γ_i} contains a product $a \diamond u$ or $u \diamond a$. Now, $\varrho_a(a \diamond u) = \varrho_a(au - ua) = u - u = 0$. Using lemma 5, this implies that $\varrho_a(x_{\gamma_i}) = 0$ and then that $\varrho_a x_\tau = 0$, which concludes the proof. □

Example 11. Using the precedent example, we can see that

$$\varrho_4 x_{(134)(25)} = 1325 - 1352 - 1325 + 1352 - 3125 + 3152 + 3125 - 3152 = 0.$$

Remark 4. The proof of proposition 6 does not use the fact that the cycle decomposition is in standard form. This condition is indeed only necessary to prove that the outputs of the algorithm for all $\tau \in \mathcal{D}_A$ constitute a basis of H_A.

We now get to the central result in the construction of our wavelet basis: $(x_\tau)_{\tau \in \mathcal{D}_A}$ is a basis of H_A for all $A \in \mathcal{P}([\![n]\!])$. In [33], they prove that their algorithm generates a basis for the top homology space of the complex of injective words. This result cannot be directly transposed in our context because H_A is not the top homology space of the complex of injective words on A. It happens however that the proof is exactly the same as the proof of theorem 5.2 in [33] and relies on concepts introduced specifically for that purpose (namely "graph derangements" and the "collapsing map"). It is thus left to the reader.

Theorem 4. *For all $A \in \mathcal{P}([\![n]\!])$, $(x_\tau)_{\mathcal{D}_A}$ is a basis of H_A.*

We are now able to construct the wavelet basis of $L(\mathfrak{S}_n)$, using the embedding operator ϕ_n. Notice that for all $\tau \in \mathfrak{S}_n \setminus \{id\}$, $\mathrm{supp}(\tau) \in \mathcal{P}([\![n]\!])$. We thus set $\psi_{id} = \psi_0 = \mathbb{1}_{\mathfrak{S}_n}$ and for $\tau \in \mathfrak{S}_n \setminus \{id\}$, we define

$$\psi_\tau = \phi_n(x_\tau) = \sum_{\pi \in \mathrm{supp}(\tau)} x_\tau(\pi) \mathbb{1}_{\mathfrak{S}_n[\pi]}. \tag{20}$$

By theorem 2, ϕ_n is an isomorphism between H_A and W_A for all $A \in \mathcal{P}([\![n]\!])$. Combined with theorem 4 we immediately obtain the following theorem.

Theorem 5. *For all $A \in \mathcal{P}([\![n]\!])$, $(\psi_\tau)_{\mathcal{D}_A}$ is a basis of W_A, and*

$$(\psi_\tau)_{\tau \in \mathfrak{S}_n} \text{ is a basis of } L(\mathfrak{S}_n).$$

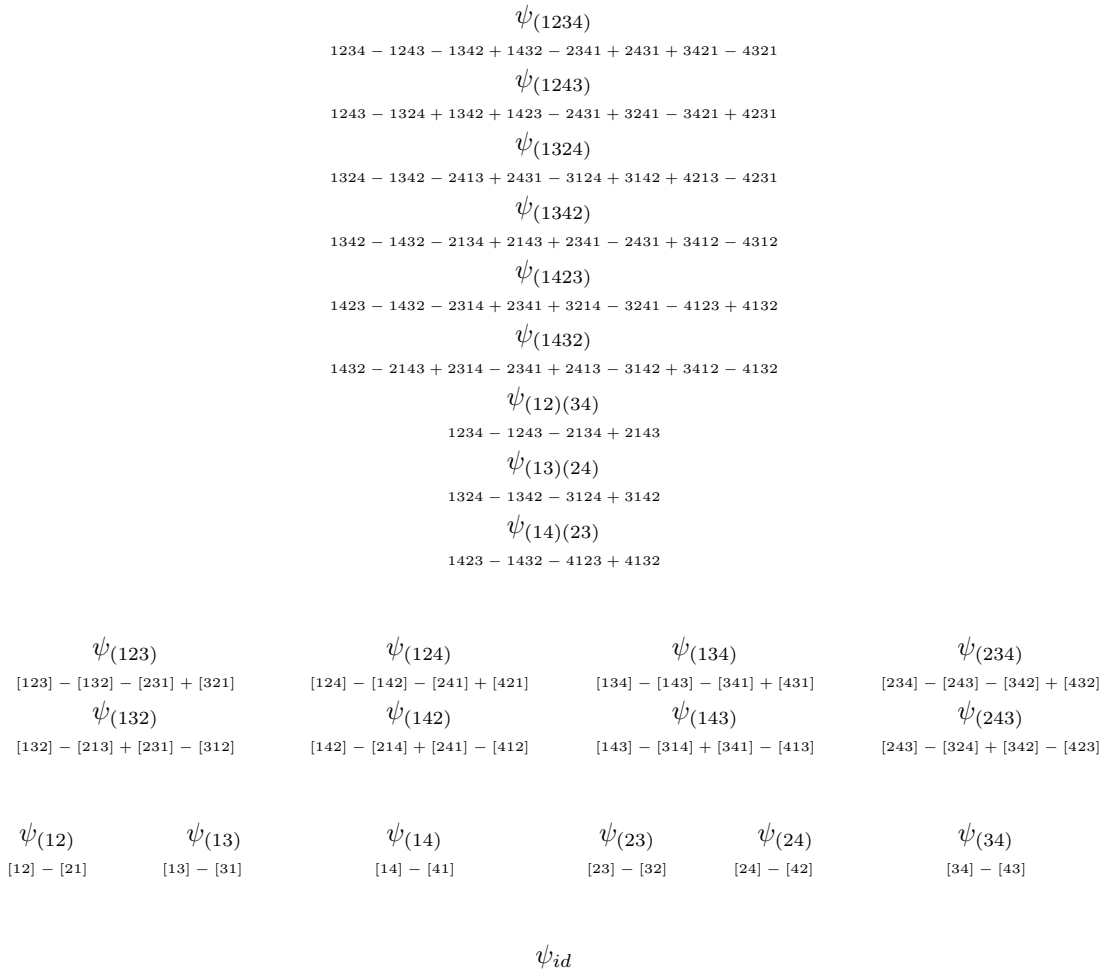


Figure 9: Wavelet basis of $L(\mathfrak{S}_4)$

Example 12. For $n = 4$, figure 9 gives the full wavelet basis of $L(\mathfrak{S}_4)$ ($[\pi]$ is a shortcut for $\mathbb{1}_{\mathfrak{S}_4[\pi]}$).

Remark 5. The wavelet basis and the multiresolution decomposition are not orthogonal, example 12 provides many couples $\tau, \tau' \in \mathfrak{S}_4$ such that $\langle \psi_\tau, \psi_{\tau'} \rangle \neq 0$.

4.3 General properties of the wavelet basis

For a chain $x \in L(\Gamma_n)$ (in particular a function in $L(\mathfrak{S}_n)$), we define its support by $\text{supp}(x) = \{\pi \in \Gamma_n \mid x(\pi) \neq 0\}$. See the Appendix section for the proof of the following proposition.

Proposition 7. *Let $\tau \in \mathfrak{S}_n \setminus \{id\}$, $k = |\tau|$ and $r = \text{cyc}(\tau)$.*

1. *$\psi_\tau(\sigma) \in \{-1, 0, 1\}$ for all $\sigma \in \mathfrak{S}_n$.*
2. *$|\text{supp}(\psi_\tau)| = 2^{k-r}(n-k+1)!$.*

This first proposition provides some general intuition about the wavelet basis. In particular, property *1.* is interesting because it means that all the properties of a wavelet function simply depend on the sign of its values and on the combinatorial structure of its support. The following proposition shows the interaction between wavelets and translations. It appears clearly in the representation of the full wavelet basis of $L(\mathfrak{S}_4)$ in example 12 that at scale k, wavelet functions in W_A with $A \in \binom{[\![n]\!]}{k}$ are the translated of wavelet functions in $W_{\{1,\dots,k\}}$. And indeed, as $(\psi_\tau)_{\tau \in \mathcal{D}_{\{1,\dots,k\}}}$ is a basis of $W_{\{1,\dots,k\}}$ (by theorem 5), $(T_{\sigma_0}\psi_\tau)_{\tau \in \mathcal{D}_{\{1,\dots,k\}}}$ is a basis of W_A for any $\sigma_0 \in \mathfrak{S}_n$ such that $\sigma_0(\{1,\dots,k\}) = A$, by proposition 4. The following proposition refines this result.

Proposition 8. *Let $\tau \in \mathfrak{S}_n$ and $\sigma_0 \in \mathfrak{S}_n$ a permutation that preserves the order of the elements of* $\mathrm{supp}(\tau)$*, i.e. if* $\mathrm{supp}(\tau) = \{a_1, \dots, a_k\}$ *with $a_1 < \dots < a_k$, then $\sigma_0(a_1) < \dots < \sigma_0(a_k)$. Then we have*

$$T_{\sigma_0}\psi_\tau = \psi_{\sigma_0 \tau \sigma_0^{-1}}.$$

Proof. If $\tau = id$, ψ_{id} is invariant under translations and the equality is trivially verified. We assume $\tau \neq id$, thus $\psi_\tau = \phi_n\, x_\tau$. By lemma 4, $T_{\sigma_0}\phi_n\, x_\tau = \phi_n\, T_{\sigma_0} x_\tau$. Let $\gamma_1 \dots \gamma_r$ be the standard cycle form of τ with $\gamma_i = (a_{i,1} \dots a_{i,k_i})$. Then it is easy to see that $T_{\sigma_0}x_\tau$ is the output of algorithm 1 when taking as input the permutation with cycle form $\gamma'_1 \dots \gamma'_r$ where $\gamma'_i = (\sigma_0(a_{i,1}) \dots \sigma_0(a_{i,k_i}))$. The order-preserving condition on σ_0 assures that this is a standard cycle form. The proof is concluded by a classic result (or a simple verification) that says that this is the cycle form of the permutation $\sigma_0 \tau \sigma_0^{-1}$. □

The third general property concerns the marginals of the wavelet functions. It actually only relies on the embedding operator ϕ_n, and not on algorithm 1. For $A \in \mathcal{P}([\![n]\!])$, we define the embedding operator $\phi_A : \bigoplus_{B \subset A} L(\Gamma(B)) \to L(\mathfrak{S}'_A)$ on the Dirac functions by

$$\phi_A : \pi \mapsto \mathbb{1}_{\mathfrak{S}'_A[\pi]} = \sum_{\sigma \in \mathfrak{S}'_A[\pi]} \sigma. \tag{21}$$

Proposition 9. *Let $A \in \mathcal{P}([\![n]\!])$ and $\pi \in \Gamma_n$ such that $c(\pi) \subset A$. Then*

$$M_A \phi_n \pi = \frac{(n - |\pi| + 1)!}{(|A| - |\pi| + 1)!}\ \phi_A \pi.$$

Proposition 9 is a direct consequence of the following lemma, of which proof is given in the Appendix section.

Lemma 6. *Let $(\pi, \pi') \in (\Gamma_n)^2$, and π^0 be the subword of π with content $c(\pi) \cap c(\pi')$. Denoting by $|\pi| = k$, $|\pi'| = l$ and $|c(\pi) \cap c(\pi')| = m$, we have*

$$|\mathfrak{S}_n[\pi] \cap \mathfrak{S}_n(\pi')| = \begin{cases} \dfrac{(n-k+1)!}{(l-m+1)!} & \text{if } \pi^0 \text{ is a contiguous subword of } \pi', \\ 0 & \text{otherwise.} \end{cases}$$

The combination of proposition 2 and 9 give an explicit formula for the marginals of any elements of a space W_B, in particular for the marginals of wavelet functions.

Proposition 10 (Marginals of the wavelet functions)**.** *Let $A \in \mathcal{P}([\![n]\!])$. $M_A\psi_{id}$ is the constant function on $\mathfrak{S}'_A$ equal to $n!/|A|!$, and for $\tau \in \mathfrak{S}_n \setminus \{id\}$,*

$$M_A\psi_\tau = \begin{cases} \dfrac{(n-|\tau|+1)!}{(|A|-|\tau|+1)!}\ \phi_A(x_\tau) & \text{if } \mathrm{supp}(\tau) \subset A, \\ 0 & \text{otherwise.} \end{cases}$$

This last proposition provides the explicit wavelet basis for the space $\mathbb{M}_{\mathcal{A}}$ for any observation design $\mathcal{A} \subset \mathcal{P}([\![n]\!])$.

Example 13. We come back to the same example as in subsection 2.5: $n = 4$ and $\mathcal{A} = \{\{1,3\},\{2,4\},\{3,4\},\{1,2,3\},\{1,3,4\}\}$. The space $\mathbb{M}_{\mathcal{A}}$ has dimension 11 and its basis is represented by figure 10 (only the marginals on subsets $A \in \mathcal{A}$ are represented, all with the same scale).

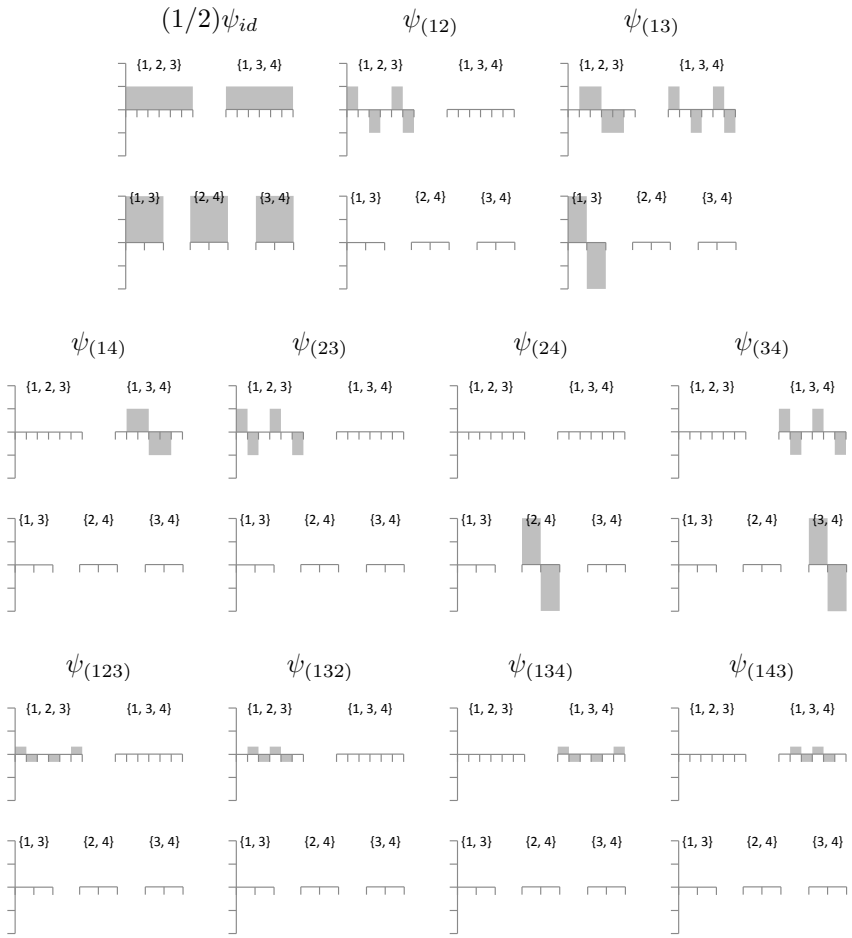


Figure 10: Wavelet basis of $\mathbb{M}_{\mathcal{A}}$, for $\mathcal{A} = \{\{1,3\},\{2,4\},\{3,4\},\{1,2,3\},\{1,3,4\}\}$

4.4 Structure of the wavelet basis

The properties of a multiresolution analysis $(\tilde{V}^j)_{j\in\mathbb{Z}}$ on $L^2(\mathbb{R})$ directly lead to the definition of an adapted wavelet basis $(\tilde{\psi}_{j,n})_{(j,n)\in\mathbb{Z}^2}$: take $\tilde{\psi}\in\tilde{V}^0$ and define $\tilde{\psi}_{j,n}(x)=2^{j/2}\tilde{\psi}(2^jx-n)$. Then $(\tilde{\psi}_{j,n})_{(j,n)\in\mathbb{Z}^2}$ is a basis of $L^2(\mathbb{R})$ adapted to $(\tilde{V}^j)_{j\in\mathbb{Z}}$ (see [26]) and has a simple interpretation, $\tilde{\psi}$ is the "mother" wavelet and all the wavelet functions are obtained from $\tilde{\psi}$ by dilation and translation. More specifically, at scale j, the wavelet function $\tilde{\psi}_{j,0}$ is obtained by dilation of $\tilde{\psi}_{j-1,0}$, $\tilde{\psi}_{j,0}(x)=\sqrt{2}\tilde{\psi}_{j-1,0}(2x)$, and all the $\tilde{\psi}_{j,n}$'s by translation of $\tilde{\psi}_{j,0}$, $\tilde{\psi}_{j,n}(x)=\tilde{\psi}_{j,0}(x-2^{-j}n)$. These relations encode the structure of the basis and are at the core of many applications.

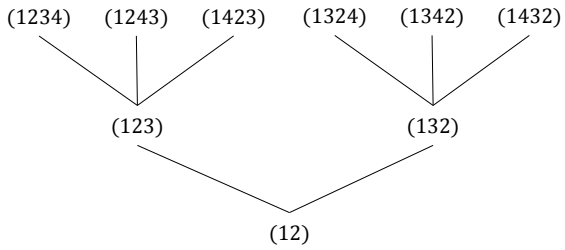
In the present setting, while the translation operators are adapted to the multiresolution decomposition, the latter can only be equipped with a dezooming operator, as explained in subsection 3.4. Hence, there is no natural operation that, in conjunction with translations, would fully encode the structure of any wavelet basis associated to it. Our wavelet basis possesses however a particular structure that stems from its generative algorithm. It is encoded in two relations that show how to obtain a wavelet chain x_τ from the wavelet chains $x_{\tau'}$ of lower scales. The first relation encode the links between wavelet chains indexed by one cycle. It uses a recursive structure on cycles, given by the following lemma, the proof of which is only technical and left in appendix.

Lemma 7. *Let $A=\{a_1,\dots,a_k\}\subset[\![n]\!]$ with $k\in\{1,\dots,n-1\}$ and $b\notin A$. Then*

1. $(a_1\ \dots\ a_k)(a_j\ b)=(a_1\ \dots\ a_j\ b\,a_{j+1}\ \dots\ a_k)\quad$ *for $j\in\{1,\dots,k\}$,*
2. $\mathrm{Cycle}(A\cup\{b\})=\{\gamma\cdot(a_j\ b)\mid j\in\{1,\dots,k\},\gamma\in\mathrm{Cycle}(A)\}$.

This lemma means that the set of cycles with support $A\cup\{b\}$ can be obtained recursively from the cycles with support A by inserting b in each cycle $\gamma\in\mathrm{Cycle}(A)$ to the right of an element a_j of this cycle. This can be represented by a tree.

Example 14. Cycles with support $\{1,2,3,4\}$ are obtained via the following tree.



For $a,b\in[\![n]\!]$, we define the elementary chain $\varepsilon_{b,a}\in L(\Gamma_n)$ by

$$\varepsilon_{b,a}(\pi)=\begin{cases}1 & \text{if } a,b\in c(\pi) \text{ and } \pi(b)-\pi(a)=1,\\ -1 & \text{if } a,b\in c(\pi) \text{ and } \pi(b)-\pi(a)=-1,\\ 0 & \text{otherwise.}\end{cases}$$

Theorem 6. *Let $\gamma=(a_1\ \dots\ a_k)$, $A=\mathrm{supp}(\gamma)$, $b>\max A$ and $j\in\{1,\dots,k\}$. Then for all $\pi\in\mathfrak{S}'_{A\cup\{b\}}$,*

$$x_{\gamma\cdot(a_j\ b)}=\varepsilon_{b,a_j}(\pi)\,x_\gamma(\pi\setminus\{b\}).$$

Proof. By lemma 7, $\gamma \cdot (a_j\, b) = (a_1 \dots a_j\, b\, a_{j+1} \dots a_k)$. Since $b > \max A$, applying algorithm 1 to $\gamma \cdot (a_j\, b)$ gives

$$\begin{aligned} x_{\gamma\cdot(a_j\, b)} &= x_{(a_1 \dots a_j\, b\, a_{j+1} \dots a_k)} \\ &= \dots \diamond \dots (a_j \diamond b) \dots \diamond \dots \\ &= \dots \diamond \dots (a_j b - b a_j) \dots \diamond \dots \\ &= \dots \diamond \dots (a_j b) \dots \diamond \dots - \dots \diamond \dots (b a_j) \dots \diamond \dots \end{aligned}$$

Thus for $\pi \in \mathfrak{S}'_{A\cup\{b\}}$,

$$\begin{aligned} x_{\gamma\cdot(a_j\, b)}(\pi) &= \begin{cases} x_\gamma(\pi \setminus \{b\}) & \text{if } \pi(b) - \pi(a_j) = 1 \\ -x_\gamma(\pi \setminus \{b\}) & \text{if } \pi(b) - \pi(a_j) = -1 \\ 0 & \text{otherwise} \end{cases} \\ &= \varepsilon_{b,a_j}(\pi)\, x_\gamma(\pi \setminus \{b\}). \end{aligned}$$

□

Example 15. For $A = \{1, 2, 3, 4\}$, for all $\pi \in \mathfrak{S}'_A$,

$$x_{(1342)}(\pi) = \varepsilon_{4,3}(\pi)\, x_{(132)}(\pi \setminus \{4\}).$$

Theorem 6 leads to an explicit formula for $x_\gamma(\pi)$ by a simple induction. It just requires some more notations. For $A \in \mathcal{P}([\![n]\!])$, we define a sequence of subsets by $A^{(0)} = A$ and

$$A^{(j)} = A^{(j-1)} \setminus \{\max A^{(j-1)}\} \qquad \text{for } j \in \{1, \dots, |A| - 1\}.$$

If $A = \{a_1, \dots, a_k\}$ with $a_1 < \dots < a_k$, then $A^{(j)} = \{a_1, \dots, a_{k-j}\}$. It is easy to see that for any $\gamma \in \mathrm{Cycle}(A)$, there exists a unique $(u_1, \dots, u_{k-1}) \in A^{(k-1)} \times \dots \times A^{(1)}$, denoted by $\mathbf{u}(\gamma)$, such that

$$\gamma = (u_1\, a_2)(u_2\, a_3) \dots (u_{k-1}\, a_k)$$

(It is given by $u_{k-1} = \gamma^{-1}(a_k)$, and $u_i = \left[\gamma\, (u_{k-1}\, a_k) \dots (u_{i+1}\, a_{i+2})\right]^{-1} (a_{i+1})$, for $i \in \{1, \dots, k-2\}$).

Corollary 2. *Let $A = \{a_1, \dots, a_k\} \subset [\![n]\!]$ with $a_1 < \dots < a_k$, and $\gamma \in \mathrm{Cycle}(A)$. We set $\mathbf{u}(\gamma) = (u_1, \dots, u_{k-1})$. Then for all $\pi \in \mathfrak{S}'_A$,*

$$x_\gamma(\pi) = \prod_{j=0}^{k-2} \varepsilon_{a_{k-j}, u_{k-j-1}}(\pi_{|A^{(j)}}).$$

Example 16. For $A = \{1, 2, 3, 4\}$, for all $\pi \in \mathfrak{S}'_A$,

$$x_{(1342)}(\pi) = \varepsilon_{4,3}(\pi)\varepsilon_{3,1}(\pi_{|\{1,2,3\}})\varepsilon_{2,1}(\pi_{|\{1,2\}}).$$

The second relation that encodes the structure of the wavelet chains gives the link between a wavelet chain indexed by a product of cycles and the wavelet chains indexed by these cycles. It stems directly from the definition of algorithm 1.

Theorem 7. *Let $\tau = \gamma_1 \dots \gamma_r \in \mathfrak{S}_n$ written in standard cycle form. Then $x_{\gamma_1 \dots \gamma_r}$ is the concatenation of $x_{\gamma_1}, \dots, x_{\gamma_r}$:*

$$x_{\gamma_1 \dots \gamma_r} = x_{\gamma_1} \dots x_{\gamma_r}.$$

For $\tau = \gamma_1 \dots \gamma_r \in \mathfrak{S}_n$ written in standard cycle form, we define the decomposition of a word $\pi \in \mathfrak{S}'_{\mathrm{supp}(\tau)}$ associated to the cycle structure of τ by the tuple of contiguous subwords $(\pi^1, \dots, \pi^r)$ such that $\pi = \pi^1 \dots \pi^r$ and $|\pi^i| = l(\gamma_i)$ for all $i \in \{1, \dots, r\}$. The explicit version of theorem 7 is given by the following corollary.

Corollary 3. *Let $\tau = \gamma_1 \dots \gamma_r \in \mathfrak{S}_n$ written in standard cycle form, and $\pi \in \mathfrak{S}'_{\mathrm{supp}(\tau)}$. Let $\left(\pi^1, \dots, \pi^r\right)$ be the decomposition of π associated to the cycle structure of τ. Then*

$$x_{\gamma_1 \dots \gamma_r}(\pi) = \begin{cases} \prod_{i=1}^{r} x_{\gamma_i}\left(\pi^i\right) & \text{if } c(\pi^i) = \mathrm{supp}(\gamma_i) \text{ for all } i \in \{1, \dots, r\}, \\ 0 & \text{otherwise.} \end{cases}$$

Example 17. Let $\tau = (134)(25) = \gamma_1\gamma_2$. We have $\mathrm{supp}(\gamma_1) = \{1, 3, 4\}$ and $\mathrm{supp}(\gamma_2) = \{2, 5\}$. The decomposition of a word $\pi = \pi_1 \dots \pi_5 \in \mathfrak{S}_5$ associated to the cycle structure of τ is given by $\pi^1 = \pi_1\pi_2\pi_3$ and $\pi^2 = \pi_4\pi_5$.

- For $\pi = 24351$, $(c(\pi^1), c(\pi^2)) = (\{2, 3, 4\}, \{1, 5\}) \neq (\mathrm{supp}(\gamma_1), \mathrm{supp}(\gamma_2))$, so
$$x_{(134)(25)}(24351) = 0.$$
- For $\pi = 41352$, $(c(\pi^1), c(\pi^2)) = (\{1, 3, 4\}, \{2, 5\}) = (\mathrm{supp}(\gamma_1), \mathrm{supp}(\gamma_2))$, so
$$x_{(134)(25)}(41352) = x_{(134)}(413)x_{(25)}(52).$$

The two relations given by theorems 6 and 7 encode the full structure of the wavelet basis, and allow to compute recursively any wavelet chain, from the wavelet chains of scale 2. We do not have an analogous concept of the "mother" wavelet in our case because the operations involved in the computation of a wavelet chain vary at each stage, but these relations remain the base for many applications, such as the design of fast decomposition algorithms in the wavelet basis.

5 Conclusion and perspectives

Exploiting the powerful formalism of injective words, we developed the first general framework to perform data analysis on incomplete rankings in the present paper. Its cornerstone is the multiresolution decomposition of $L(\mathfrak{S}_n)$ in function of the spaces W_A, that provides a decomposition of the space $\mathbb{M}_{\mathcal{A}}$ for any observation design $\mathcal{A}$. The explicit wavelet basis Ψ adapted to this multiresolution decomposition is the key to use this framework in practice, allowing to perform linear or nonlinear approximation in any space $\mathbb{M}_{\mathcal{A}}$. It paves the way for many statistical applications, such as estimation of a ranking distribution or prediction of a ranking on a new subset of items, aggregation of many incomplete rankings into one full ranking or clustering of incomplete rankings. All these applications require the design of fast decomposition algorithms as well as the theoretical study of the properties of the wavelet basis regarding (nonlinear) approximation. This will be the subject of forthcoming articles. At last, another line of further research consists in trying to generalize the present framework to incomplete rankings which also allow ties.

References

[1] Hossein Azari Soufiani, William Chen, David C Parkes, and Lirong Xia. Generalized method-of-moments for rank aggregation. In *Advances in Neural Information Processing Systems 26*, pages 2706–2714, 2013.

[2] J.P. Barthélémy and B.Montjardet. The median procedure in cluster analysis and social choice theory. *Mathematical Social Sciences*, 1:235–267, 1981.

[3] A. Björner and M. L. Wachs. On lexicographically shellable posets. *Trans. Amer. Math. Soc.*, 277:323–341, 1983.

[4] R.R. Coifman and M. Maggioni. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 21:53–94, 2006.

[5] Onkar Dalal, Srinivasan H. Sengemedu, and Subhajit Sanyal. Multi-objective ranking of comments on web. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 419–428, 2012.

[6] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes - Monograph Series. Institute of Mathematical Statistics, Hayward, CA, 1988.

[7] Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, 26(1):363–397, 1998.

[8] F.D. Farmer. Cellular homology for posets. *Math. Japon*, 23:607–613, 1978/79.

[9] Y. Freund, R. D. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *JMLR*, 4:933–969, 2003.

[10] Matan Gavish, Boaz Nadler, and Ronald R. Coifman. Multiscale wavelets on trees, graphs and high dimensional data: theory and applications to semi supervised learning. In *International Conference on Machine Learning*, pages 567–574, 2010.

[11] David K. Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129 – 150, 2011.

[12] Phil Hanlon and Patricia Hersh. A Hodge decomposition for the complex of injective words. *Pacific J. Math.*, 214(1):109–125, 2004.

[13] David P. Helmbold and Manfred K. Warmuth. Learning permutations with exponential weights. *Journal of Machine Learning Research*, 10:1705–1736, 2009.

[14] J. Huang and C. Guestrin. Riffled independence for ranked data. In *Proceedings of NIPS'09*, 2009.

[15] J. Huang, C. Guestrin, and L. Guibas. Fourier theoretic probabilistic inference over permutations. *JMLR*, 10:997–1070, 2009.

[16] E. Hüllermeier, J. Fürnkranz, W. Cheng, and K. Brinker. Label ranking by learning pairwise preferences. *Artificial Intelligence*, 172:1897–1917, 2008.

[17] David R. Hunter. MM algorithms for generalized Bradley-Terry models. *The Annals of Statistics*, 32:384–406, 2004.

[18] Ekhine Irurozki, Borja Calvo, and J Lozano. Learning probability distributions over permutations by means of Fourier coefficients. *Advances in Artificial Intelligence*, pages 186–191, 2011.

[19] Srikanth Jagabathula and Devavrat Shah. Inferring Rankings Using Constrained Sensing. *IEEE Transactions on Information Theory*, 57(11):7288–7306, 2011.

[20] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. *Math. Program.*, 127(1):203–244, 2011.

[21] Ramakrishna Kakarala. A signal processing approach to Fourier analysis of ranking data: the importance of phase. *IEEE Transactions on Signal Processing*, pages 1–10, 2011.

[22] Risi Kondor and Marconi S. Barbosa. Ranking with kernels in Fourier space. In *COLT*, pages 451–463, 2010.

[23] Risi Kondor and Walter Dempsey. Multiresolution analysis on the symmetric group. In *Neural Information Processing Systems 25*, 2012.

[24] G. Lebanon and Y. Mao. Non-parametric modeling of partially ranked data. *JMLR*, 9:2401–2429, 2008.

[25] R. D. Luce. *Individual Choice Behavior*. Wiley, 1959.

[26] Stéphane Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE*, II(7), 1989.

[27] Stéphane Mallat. Multiresolution appxorimations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Transactions of the AMS*, 315(1), 1989.

[28] B. Mandhani and M. Meila. Tractable search for learning exponential models of rankings. In *Proceedings of AISTATS'09*, 2009.

[29] J. I. Marden. *Analyzing and Modeling Rank Data*. CRC Press, London, 1996.

[30] Y. Meyer. *Wavelets and operators:Advanced Mathematics*. Cambridge University Press, 1992.

[31] B. Osting, C. Brune, and S. Osher. Enhanced statistical rankings via targeted data collection. In *Journal of Machine Learning Research, W&CP (ICML 2013)*, volume 28 (1), pages 489–497, 2013.

[32] R. L. Plackett. The analysis of permutations. *Applied Statistics*, 2(24):193–202, 1975.

[33] Kári Ragnarsson and Bridget Eileen Tenner. Homology of the boolean complex. *Journal of Algebraic Combinatorics*, 34(4):617–639, 2011.

[34] Idan Ram, Michael Elad, and Israel Cohen. Generalized tree-based wavelet transform. *IEEE Transactions on Signal Processing*, 59(9):4199–4209, 2011.

[35] Victor Reiner, Franco Saliola, and Volkmar Welker. Spectra of symmetrized shuffling operators. *Memoirs of the American Mathematical Society*, 228(1072), 2013.

[36] Victor Reiner and Peter Webb. Combinatorics of the bar resolution in group cohomology. *J. Pure Appl. Algebra*, 190:291–327, 2004.

[37] Dan Rockmore, Peter Kostelec, Wim Hordijk, and Peter F. Stadler. Fast Fourier transform for fitness landscapes. *Applied and Computational Harmonic Analysis*, 12(1):57–76, January 2002.

[38] Raif M. Rustamov. Average interpolating wavelets on point clouds and graphs. *CoRR*, abs/1110.2227, 2011.

[39] Mingxuan Sun, Guy Lebanon, and Paul Kidwell. Estimating probabilities in recommendation systems. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):471–492, 2012.

6 Appendix

6.1 Background on group theory

A group is a set G equipped with an associative operation $G^2 \to G, (g,h) \mapsto gh$ and an element $e \in G$ such that for all $g \in G$, $ge = eg = g$ and there exists $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$. The element e is called the identity element, and g^{-1}, necessarily unique, is called the inverse of $g \in G$. The operation is not necessarily commutative. A subgroup of G is a subset $H \subset G$ such that $e \in H$ and for all $(h,h') \in H^2$, $hh' \in H$. A left coset of a subgroup H of G is a subset (usually not a subgroup) of G of the form $\{gh \mid h \in H\}$ with $g \in G$. A simple result states that for any subgroup H of a finite group G, all the left cosets of H have same cardinality $|H|$ and they constitute a partition of G.

An action of a group G over a set E is an operation $G \times E \to E, (g,x) \mapsto g \cdot x$ such that for all $(g,g') \in G^2$ and $x \in E$, $e \cdot x = x$ and $g' \cdot (g \cdot x) = (g'g) \cdot x$. For $x \in E$, its orbit under the action of G is the set $O_x = \{g \cdot x \mid g \in G\}$, and its stabilizer is the subgroup of G $\{g \in G \mid g \cdot x = x\}$. A subset of E is an orbit of G if it is equal to a O_x. The collection of all the orbits of G is a partition of E. The action of G on E is called transitive if it has only one orbit (E), *i.e.* if for all $x \in E$, $O_x = E$.

A representation of a group G is couple (V, ρ) where V is a linear space and ρ a mapping $\rho : G \to GL(V)$, where $GL(V)$ is the group of invertible linear maps from V to V, such that for all $(g,g') \in G^2$, $\rho(gg') = \rho(g)\rho(g')$. We speak indifferently of the representation (V, ρ), the representation ρ or the representation V. When G acts transitively on a finite set E, there is a canonical representation of G on $L(E)$, called the permutation representation, defined on the Dirac functions by $\rho(g)\delta_x = \delta_{g \cdot x}$, for $x \in E$. From an analytical point of view, the operators $\rho(g)$ are exactly the translations operators on $L(E)$ associated to the action of G, and besides, for all $f \in L(E)$, $g \in G$ and $x \in E$, $(\rho(g)f)(x) = f(g^{-1} \cdot x)$. When $E = G$, this representation is called the regular representation.

A representation (V, ρ) of G is called irreducible if $V \neq \{0\}$ and there is no subspace $W \subset V$ such that $\rho(g)(W) \subset W$ for all $g \in G$ other than $\{0\}$ and V. Two representations (V_1, ρ_1) and (V_2, ρ_2) of a group G are isomorphic if there exists an isomorphism ϕ between V_1 and V_2 such that $\phi(\rho_1(g)v) = \rho_2(g)\phi(v)$ for all $g \in G$ and $v \in V$. Irreducible representations of a group are assimilated to their equivalence class of isomorphic representations.

A major result in the representation theory of finite groups is that the number of irreducible representations of a finite group G is finite (actually equal to the number of conjugacy classes of G) and that any finite-dimensional representation V of G admits a decomposition as a direct sum of irreducible representations. The number of copies of one irreducible representation in this decomposition is called its multiplicity. The decomposition of the regular representation $L(G)$ involves all the irreducible representations of G, each appearing with multiplicity equal to its dimension. If $\mathrm{Irr}(G)$ denotes the set of irreducible representations of G, then

$$L(G) \cong \bigoplus_{W \in \mathrm{Irr}(G)} d_W W,$$

where for $W \in \mathrm{Irr}(G)$, $d_W = \dim W$. See [6] for more developments on group representation theory.

6.2 Technical proofs

Proof of lemma 1. Let $(A, B) \in \mathcal{P}([\![n]\!])^2$ with $A \subset B$. The permutation group $\mathfrak{S}_A$ acts on $\mathfrak{S}'_A$ and $\mathfrak{S}'_B$. The mapping $r_{B,A} : \mathfrak{S}'_B \to \mathfrak{S}'_A, \sigma \mapsto \sigma_{|A}$, is equivariant for this action, *i.e.*, for any $\tau \in \mathfrak{S}_A$ and $\sigma \in \mathfrak{S}'_B$, $r_{B,A}(\tau \cdot \sigma) = \tau \cdot r_{B,A}(\sigma)$. The action being transitive on $\mathfrak{S}'_A$, $r_{B,A}$ is surjective. Moreover, for $\pi \in \mathfrak{S}'_A$, $\mathfrak{S}'_B(\tau \cdot \pi) = r^{-1}_{B,A}(\{\tau \cdot \pi\}) = \tau \cdot r^{-1}_{B,A}(\{\pi\}) = \tau \cdot \mathfrak{S}'_B(\pi)$. Consequently $|\mathfrak{S}'_B(\pi)| = |\mathfrak{S}'_B(\sigma \cdot \pi)|$, which, combined with $\mathfrak{S}'_B = \sqcup_{\pi \in \mathfrak{S}'_A} r^{-1}_{B,A}(\pi)$, gives the sought result. □

Proof of lemma 3. Let $\omega \in \Gamma_n$, $a \in [\![n]\!] \setminus c(\omega)$ and $\pi \in \Gamma_n$. If $c(\pi) \cap c(\omega) \neq \emptyset$, then also $c(\varrho_a \pi) \cap c(\omega) \neq \emptyset$, and both $\mathfrak{i}_\omega \pi$ and $\mathfrak{i}_\omega \varrho_a \pi$ are equal to 0 by definition. If $c(\pi) \cap c(\omega) = \emptyset$, then $\mathfrak{i}_\omega \pi = \omega \pi$. Since $a \notin c(\omega)$, it can only be deleted in the word $\omega\pi$ if it is deleted from π. This means that $\varrho_a \omega \pi = \omega \varrho_a \pi$, whether $a \in c(\pi)$ or not. We prove identically that $\varrho_a \mathfrak{j}_\omega = \mathfrak{j}_\omega \varrho_a$. □

Proof of proposition 7. The proof of this proposition is a simple analysis of algorithm 1. For a cycle $\gamma = (a_1 \dots a_k)$, the associated x_γ is equal to an expression of the form $a_1 \diamond \cdots \diamond a_k$ with a particular way to put parentheses. When expanded, this expression gives 2^{k-1} terms with sign $+$ or $-$ between them. It could happen that some of the terms are the same and thus add or balance. But actually, for $x \in L(\Gamma(A))$ with $A \subset [\![n]\!]$, $1 \leq |A| \leq n-1$ and $b \in [\![n]\!] \setminus A$, $\mathrm{supp}(x \diamond b) = \{\pi b \mid \pi \in \mathrm{supp}(x)\} \sqcup \{b\pi \mid \pi \in \mathrm{supp}(x)\}$. By recursion, we obtain that $|\mathrm{supp}(x_\gamma)| = 2^{k-1}$, meaning also that all the terms in the expanded version of $a_1 \diamond \cdots \diamond a_k$ are different. Furthermore, for $x \in L(\Gamma(A))$ and $y \in L(\Gamma(B))$ with $A, B \subset [\![n]\!]$, $A, B \neq \emptyset$ and $A \cap B = \emptyset$, we have $|\mathrm{supp}(xy)| = |\mathrm{supp}(x)||\mathrm{supp}(y)|$. Now, let $\tau = \gamma_1 \dots \gamma_r$ be a permutation written in standard cycle form, with $\gamma_i = (a_{i,1} \dots a_{i,k_i})$. Then $x_\tau = (a_{1,1} \diamond \cdots \diamond a_{1,k_1}) \dots (a_{r,1} \diamond \cdots \diamond a_{r,k_r})$, and this expression expands in $2^{k_1 - 1} \dots 2^{k_r - 1} = 2^{k-r}$ different terms. This shows both that $|\mathrm{supp}(x_\tau)| = 2^{k-r}$ and that x_τ takes its values in $\{-1, 0, 1\}$. Applying ϕ_n concludes the proof. □

Proof of lemma 6. Let $(\pi, \pi') \in (\Gamma_n)^2$, and π^0 be the subword of π with content $c(\pi) \cap c(\pi')$. We denote by $|\pi| = k$, $|\pi'| = l$ and $|c(\pi) \cap c(\pi')| = m$. By definition, $\mathfrak{S}_n[\pi] \cap \mathfrak{S}_n(\pi') = \{\sigma \in \mathfrak{S}_n \mid \sigma$ admits π as a contiguous subword and π' as a subword$\}$. If π^0 is not a contiguous subword of π', then there exist a subword π^* of π^0 which is a contiguous subword of π, $a \in c(\pi') \setminus c(\pi^0)$ and $i \in \{2, \dots, m\}$ such that $\pi^* \triangleleft_i a$ is a subword of π'. So if $\sigma \in \mathfrak{S}_n[\pi] \cap \mathfrak{S}_n(\pi')$, σ admits *a fortiori* π^* as a contiguous subword and $\pi^* \triangleleft_i a$ as a subword, which is not possible. Hence, $|\mathfrak{S}_n[\pi] \cap \mathfrak{S}_n(\pi')| = 0$ in this case. We now assume that π^0 is a contiguous subword of π'. Let $i \in \{1, \dots, l\}$ such that

$\pi'_i \dots \pi'_{i+m-1} = \pi^0$. Then each element of $\mathfrak{S}_n[\pi] \cap \mathfrak{S}_n(\pi')$ can be seen as a way of filling the blanks denoted by _____ with all the elements of $[\![n]\!] \setminus (c(\pi) \cup c(\pi'))$, in the following figure.

$$\_\_\_\_\_\ \pi'_1\ \_\_\_\_\_\ \dots\ \_\_\_\_\_\ \pi'_{i-1}\ \_\_\_\_\_\ \pi\ \_\_\_\_\_\ \pi'_{i+m}\ \dots \_\_\_\_\_\ \pi'_l\ \_\_\_\_\_.$$

If we do not take the order of the elements into account, the number of such fillings is equal to the number of ways of putting $n-(k+l-m)$ indistinguishable balls (the elements of $[\![n]\!] \setminus (c(\pi) \cup c(\pi'))$) into $l-m+2$ boxes (the blanks). From a classic result in combinatorics, this number is equal to

$$\binom{(n-(k+l-m))+(l-m+2)-1}{(l-m+2)-1} = \binom{n-k+1}{l-m+1}.$$

Now, to take the order into account, we have to multiply by the number of possible reorderings of the filling elements, equal to $(n-(k+l-m))!$. The final result is thus

$$\binom{n-k+1}{l-m+1}(n-(k+l-m))! = \frac{(n-k+1)!}{(l-m+1)!}.$$

□

Proof of lemma 7. Let $A = \{a_1, \dots, a_k\} \subset [\![n]\!]$ with $k \in \{1, \dots, n-1\}$ and $b \notin A$. For $j \in \{1, \dots, k\}$, $\gamma = (a_1 \ \dots \ a_k)$ and $\tau = (a_1 \ \dots \ a_k)(a_j \, b)$, we have

$$\begin{aligned}
\tau(a_i) &= \gamma(a_i) = a_{i+1} && \text{for } i \in \{1, \dots, k\} \setminus \{j\} \text{ with } a_{k+1} = a_1 \text{ by convention,} \\
\tau(a_j) &= \gamma(b) = b, \\
\tau(b) &= \gamma(a_j) = a_{j+1}, \\
\tau(a') &= \gamma(a') = a' && \text{for all } a' \notin A \cup \{b\}.
\end{aligned}$$

Hence $\tau = (a_1 \ \dots \ a_j \, b \, a_{j+1} \ \dots \ a_k)$. This proves 1. and at the same time that $\{\gamma \cdot (a_j \, b) \mid j \in \{1, \dots, k\}, \gamma \in \mathrm{Cycle}(A)\} \subset \mathrm{Cycle}(A \cup \{b\})$. Now, let $\gamma \in \mathrm{Cycle}(A \cup \{b\})$, $a^* = \gamma^{-1}(b)$ and $\gamma' \in \mathrm{Cycle}(A)$ be the cycle obtained when deleting b in γ. Then by 1., $\gamma = \gamma' \cdot (a^* \, b)$. This concludes the proof. □