Ordinal symbolic dynamics

Karsten Keller and Mathieu Sinn
Mathematical Institute, Wallstraße 40, 23560 Lübeck, Germany

Abstract

Ordinal time series analysis is a new approach to the investigation of long and complex time series. Here 'ordinal' means to deal with the order relations between successive values instead of the values themselves. In this paper we discuss ordinal time series analysis from the symbolic dynamics viewpoint. In particular, we introduce a special transformation extracting the ordinal information from a time series and consider invariance properties of ordinal time series analysis.

1 Introduction

Ordinal patterns. The idea of ordinal time series analysis due to Bandt and Pompe [1] is to quantify the up-and-down in a given time series. This is done by considering ordinal patterns which describe the order relations between some equidistant successive values and determining the distribution of the ordinal patterns in a time series or in parts of it. Note that one detail of the distribution is given by the permutation entropy introduced by Bandt and Pompe as a measure for quantifying the complexity of a time series and the system behind it. It is relatively robust with respect to observational and dynamical noise and strongly related to the Kolmogorov-Sinai entropy in the case of time series coming from one-dimensional dynamical systems (see [1, 2]).

Here we use the concept of an ordinal pattern as it is given in [4, 6]). We assume a real-valued time series $(x_t)_{t\in T}$, where $T = \mathbb{Z}$ for theoretical considerations. In praxis $T = \{r, r+1, \ldots, s-1, s\}$ for some $r, s \in \mathbb{Z}$. All concepts will be given for $T = \mathbb{Z}$, the adaption to the practical case, however, will be obvious. Ordinal patterns will be given for positive integers τ and d called *delay* and *order*, respectively. While different delays reflect different details of a time series, with increasing order the information obtained increases. In the following let $(x_t)_{t\in\mathbb{Z}}$ be a real-valued time series and $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Definition 1. *The* ordinal pattern *of order $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ at time t is defined as the unique permutation*

$$\pi_d^\tau(t) = \begin{pmatrix} 0 & 1 & 2 & \ldots & d \\ r_0 & r_1 & r_2 & \ldots & r_d \end{pmatrix} =: (r_0, r_1, r_2, \ldots, r_d)$$

of the set $\{0, 1, \ldots, d\}$ *satisfying*

$$x_{t-r_0\tau} \geq x_{t-r_1\tau} \geq \ldots \geq x_{t-r_{d-1}\tau} \geq x_{t-r_d\tau} \tag{1}$$

and

$$r_{l-1} > r_l \text{ if } x_{t-r_{l-1}\tau} = x_{t-r_l\tau}. \tag{2}$$

Remark. For technical reasons we consider permutations of a set $\{0, 1, 2, \ldots, d\}$ instead of $\{1, 2, \ldots, d\}$ as usual.

What in fact $\pi_d^\tau(t)$ describes is how the order of the considered times

$$t - \mathbf{0}\tau > t - \mathbf{1}\tau > \ldots > t - (\mathbf{d-1})\tau > t - \mathbf{d}\tau$$

is turned into the order of the corresponding values. Let us give an example.

Example 2. Consider a part of an EEG (electroencephalogram) time series consisting of 140 points (see Figure 1). Let $t = 120, \tau = 20$ and $d = 5$. Then we have $x_{t-\mathbf{0}\tau} = x_{120} > x_{t-\mathbf{3}\tau} = x_{60} > x_{t-\mathbf{4}\tau} = x_{40} > x_{t-\mathbf{5}\tau} = x_{20} > x_{t-\mathbf{1}\tau} = x_{100} > x_{t-\mathbf{2}\tau} = x_{80}$, hence $\pi = \pi_d^\tau = (\mathbf{0}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{1}, \mathbf{2})$. As the accompaning upper curve in Figure 1 shows, setting $x_{t-r_l\tau}$ to $d - l$ for $l \in \{0, 1, \ldots, d\}$ yields a qualitative reconstruction of the original signal, preserving in particular monotone parts.

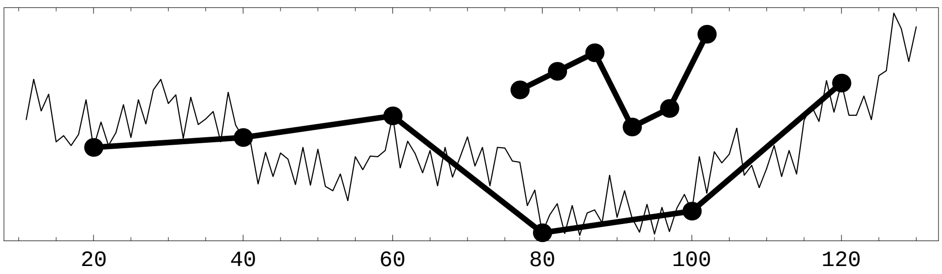

Figure 1: Ordinal pattern

As formula (1) shows, we do not emphasize equality of two values. (2) guaranties uniqueness of the permutation defined by (1), in only one of many possible ways. This point will be discussed at the end of the given section.

The symbolic dynamics viewpoint: Decompositions. It is suggestive to consider ordinal patterns from a special view: Obviously, ordinal patterns $\pi_d^\tau(t)$ only depend on the vectors $(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-d\tau}) \in \mathbb{R}^{d+1}$. In particular, according to (1) and (2)

$$\pi_d^\tau(t) = \pi_d^\tau(s) \text{ iff for all } k, l \in \{0, 1, \ldots, d\} \text{ with } r < l \text{ it holds}$$
$$x_{t-l\tau} \geq x_{t-r\tau} \Leftrightarrow x_{s-l\tau} \geq x_{s-r\tau}.$$

This decomposes the $\mathbb{R}^{d+1}$ into $(d+1)!$ pieces with the same 'ordinal pattern'.

The inverse method, i.e. dividing a state space into a few number of pieces and only using the time-dependent information in which piece a state lies, is called *symbolic dynamics*. We want to discuss symbolic dynamics for finite partitions of the $\mathbb{R}^{d+1}$ only depending on the ordinal structure of a vector. This is what we mean by *ordinal symbolic dynamics*. We say that two vectors are *separated* by a partition $\mathfrak{D} = \{D_1, D_2, \ldots, D_p\}$ of the $\mathbb{R}^{d+1}$ if they lie in different pieces D_{q_1}, D_{q_2} of $\mathfrak{D}$.

Definition 3. *Let $\mathfrak{D}_d^{ord}$ be that partition of the $\mathbb{R}^{d+1}$ which does not separate two vectors $(v_0, v_1, \ldots, v_d)$ and $(w_0, w_1, \ldots, w_d)$ iff*

$$\text{(3)} \qquad \textit{for all } k, l \in \{0, 1, \ldots, d\} : \; v_k > v_l \Leftrightarrow w_k > w_l.$$

Since either $v_k > v_l$ or $v_k < v_l$ or $v_k = v_l$, the partition defined by (3) can be described as follows, in imitation of Definition 1:
For a vector $\mathbf{v} = (v_0, v_1, \ldots, v_d) \in \mathbb{R}^{d+1}$, let

$$\boldsymbol{\pi}^*(\mathbf{v}) = (r_0, *_1, r_1, *_2, \ldots, *_{d-1}, r_{d-1}, *_d, r_d)$$

with $\{r_0, r_1, \ldots, r_d\} = \{0, 1, \ldots, d\}$ and $*_i \in \{>, =\}$ for $i = 1, 2, \ldots, d$, such that $v_{r_0} *_1 v_{r_1} *_2 \ldots *_{d-1} v_{r_{d-1}} *_d v_{r_d}$ and

$$\text{(4)} \qquad r_{l-1} > r_l \text{ if } *_l \text{ is equal to } =.$$

Two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{d+1}$ are separated by $\mathfrak{D}_d^{ord}$ iff $\boldsymbol{\pi}^*(\mathbf{v}) \neq \boldsymbol{\pi}^*(\mathbf{w})$, hence there is a one-to-one correspondence between the 'symbols' $(r_0, *_1, r_1, *_2, \ldots, *_{d-1}, r_{d-1}, *_d, r_d)$ with (4) and the pieces of $\mathfrak{D}_d^{ord}$. With

$$\boldsymbol{\pi}(\mathbf{v}) = (r_0, r_1, \ldots, r_{d-1}, r_d)$$

the relationship between this symbolization and ordinal patterns is the following:

$$\begin{aligned} \boldsymbol{\pi}^*((x_t, x_{t-\tau}, \ldots, x_{t-d\tau})) &= (r_0, *_1, r_1, *_2, \ldots, *_d, r_d) \\ &\Longrightarrow \pi_d^\tau(t) = \boldsymbol{\pi}((x_t, x_{t-\tau}, \ldots, x_{t-d\tau})) = (r_0, r_1, \ldots, r_d). \end{aligned}$$

In contrast to the ordinal patterns, $\mathfrak{D}_d^{ord}$ considers equality between one or more components of the vectors in $\mathbb{R}^{d+1}$ separately. For many applications this equality does not seem to be relevant because it occurs in an extremely few number of cases or theoretically its occurrence has probability 0. In this situation, but also in order to map the $\mathbb{R}^{d+1}$ onto the permutation group, we add pieces with equality of some components to 'nearby pieces' with no equality. Clearly, there are different ways of doing this and usually there is no best choice. We have preferred ordinal patterns because its determination is advantageous from the computational viewpoint.

Contents of the paper. On the base of ordinal patterns we define a transformation extracting the ordinal structure of a time series. It turns a given time series into a 'symbolic' time series with values in the interval $[0, 1]$ and is called the *ordinal* *transformation.* We start by focussing on aspects of representing and determining ordinal patterns in Section 2 which will lead us to a natural enumeration of ordinal patterns. Normalizing and limiting the number representations allows us to define the ordinal transformation in Section 3. Describing the arrangement of the ordinal patterns in the interval $[0,1]$ we give a characterization of the ordinal transformation from the symbolic dynamics viewpoint. In Section 4 we illustrate by examples from EEG (electroencephalography) analysis how the ordinal transformation can be applied for visualizing intrinsic structure of long and complex time series. Section 5 provides a supplementary discussion of invariance properties of ordinal symbolic dynamics.

2 Ordinal pattern representation

The way of describing ordinal patterns by permutations as in Definition 1 is plausible from the geometrical viewpoint (compare Fig. 1), we introduce however an other representation of permutations. The advantages will become evident later on.

Inversions. Consider a permutation

$$\pi = \begin{pmatrix} 0 & 1 & 2 & \dots & d \\ r_0 & r_1 & r_2 & \dots & r_d \end{pmatrix} = (r_0, r_1, r_2, \dots, r_d) \tag{5}$$

of $\{0, 1, \dots, d\}$. For $l = 1, 2, \dots, d$ let

$$i_l = i_l(\pi) = \#\{r \in \{0, 1, \dots, l-1\} \mid \pi^{-1}(r) > \pi^{-1}(l)\}$$

be the number of all $r = 0, 1, \dots, l-1$ for which the position of r is greater than the position of l. Then π is uniquely coded by the sequence $(i_1, i_2, \dots, i_d)$. Note that the sum $\sum_{l=1}^{d} i_l$ is equal to the number $\#\{(m,n) \mid m < n, \pi(n) > \pi(m)\}$ of all inversions of the permutation π. It defines a metric on the symmetric group known as *Kendal's Tau.* (For some more background see [3].) Below we will consider a 'weighted' sum of the $i_1, i_2, ..., i_d$ yielding number representations of permutations.

The inverse procedure from $(i_1, i_2, \dots, i_d)$ to π is provided via a sequence of permutations $\pi_0, \pi_1, \dots, \pi_d = \pi$ of $\{0\}, \{0,1\}, \dots, \{0, 1, \dots, d\}$ being constructed recursively:

1. $\pi_0 = (0)$ is the trivial permutation of the single set $\{0\}$.
2. When $\pi_{l-1} = (\rho_0, \rho_1, \dots, \rho_{l-1})$ for $0 < l \leq d$ is already given, π_l is obtained from π_{l-1} by inserting l into $(\rho_0, \rho_1, \dots, \rho_{l-1})$ right to ρ_{l-1} if $i_l = 0$, and left to ρ_{l-i_l} else.

Remark. For π as obtained in Example 2 it holds $i_1 = i_2 = 0$ and $i_3 = i_4 = i_5 = 2$, and the insertion process $\pi_0 \to \pi_1 \to \pi_2 \to \pi_3 \to \pi_4 \to \pi_5 = \pi$ defined by $(i_1, i_2, i_3, i_4, i_5)$ is

$$(0) \to (0,1) \to (0,1,2) \to (0,3,1,2) \to (0,3,4,1,2) \to (0,3,4,5,1,2).$$

Note that from the computational viewpoint performing the procedures between some single π and $(i_1, i_2, ..., i_d)$ is strongly related to sorting. In particular, the use of the idea of merge sorting (see [7]) provides an optimal algorithms (for large d).

Efficient coding. One easily sees that

$$i_l(\pi_d^\tau(t)) = i_l^\tau(t) = \#\{r \in \{0, 1, \ldots, l-1\} \mid x_{t-l\tau} \geq x_{t-r\tau}\}$$

for $l = 1, 2, ..., d$, hence

$$\pi_d^\tau(t) \text{ is coded by } (i_1^\tau(t), i_2^\tau(t), \ldots, i_d^\tau(t)).$$

If the permutation $\pi_d^\tau(t)$ is coded by $(i_1^\tau(t), i_2^\tau(t), \ldots, i_d^\tau(t))$, then for $l = 1, 2, \ldots, d-1$ the permutation $\pi_l^\tau(t)$ is coded by $(i_1^\tau(t), i_2^\tau(t), \ldots, i_l^\tau(t))$.

The following equality is important for an effective determination of ordinal patterns in a time series:

$$\begin{aligned} i_l^\tau(t+\tau) &= \#\{r \mid 0 \leq r < l,\, x_{t+\tau-l\tau} \geq x_{t+\tau-r\tau}\} \\ &= \#\{r \mid 0 \leq r < l,\, x_{t-(l-1)\tau} \geq x_{t-(r-1)\tau}\} \\ &= \#\{r \mid -1 \leq r < l-1,\, x_{t-(l-1)\tau} \geq x_{t-r\tau}\} \\ &= i_{l-1}^\tau(t) + i_1^{l\tau}(t+\tau) \\ &= \begin{cases} i_{l-1}^\tau(t) + 1 & \text{if } x_{t-(l-1)\tau} \geq x_{t+\tau} \\ i_{l-1}^\tau(t) & \text{else} \end{cases}. \end{aligned}$$

With this it takes d comparisons and at most d incrementation operations to compute $(i_1^\tau(t), i_2^\tau(t), \ldots, i_d^\tau(t))$ if $(i_1^\tau(t-\tau), i_2^\tau(t-\tau), \ldots, i_{d-1}^\tau(t-\tau))$ is given, which yields an efficient method for computing successive equidistant patterns.

Number representation. We want to represent a permutation π as given in (5) by a number $n_d \in \{0, 1, \ldots, (d+1)! - 1\}$. For the associated $(i_1, i_2, \ldots, i_d) \in \{0, 1\} \times \{0, 1, 2\} \times \ldots \times \{0, 1, \ldots, d\}$ and $\tilde{d} \in \mathbb{N}$ with $\tilde{d} \leq d$ let

$$n_{\tilde{d}} = \sum_{l=1}^{\tilde{d}} i_l \frac{(\tilde{d}+1)!}{(l+1)!}.$$

Here it is important to note that

$$n_{\tilde{d}} = (\tilde{d}+1) n_{\tilde{d}-1} + i_{\tilde{d}}, \tag{6}$$

with the natural definition $(i_0 =) n_0 = 0$, such that $n_{\tilde{d}-1}$ is the integer part of $\frac{n_{\tilde{d}}}{\tilde{d}+1}$ and $i_{\tilde{d}}$ the remainder of $n_{\tilde{d}}$ for division by $\tilde{d}+1$. This allows to reconstruct $i_1, i_2, \ldots, i_d$ from n_d step by step. By (6) and induction on $\tilde{d}$, one easily obtains the following statement:

Lemma 4. *For each $d \in \mathbb{N}$, the assignment*

$$(i_1, i_2, \ldots, i_d) \mapsto n_d = \sum_{l=1}^{d} i_l \frac{(d+1)!}{(l+1)!} \tag{7}$$

defines a bijection from the set $\mathcal{I}_d = \{0,1\} \times \{0,1,2\} \times \ldots \times \{0,1,\ldots,d\}$ onto $\{0,1,\ldots,(d+1)!-1\}$ which turns the lexicographic order on $\mathcal{I}_d$ into the usual order on $\{0,1,\ldots,(d+1)!-1\}$. $\square$

This yields an enumeration of permutations (ordinal patterns), on which the ordinal transformation bases. Note in particular that $n_d = 0$, $(i_1, i_2, \ldots, i_d) = (0,0,\ldots,0)$ and $\pi = (0,1,\ldots,d)$ are equivalent, and $n_d = (d+1)!-1$, $(i_1, i_2, \ldots, i_d) = (1,2,\ldots,d)$ and $\pi = (d, d-1, \ldots, 1, 0)$ too. With respect to the number representation $(0,1,\ldots,d)$ and $(d,d-1,\ldots,1,0)$ have maximal distance, which is natural since the patterns represent completely opposite monotone behavior. This and the general arrangement of ordinal patterns will be discussed in relation to the class of ordinal transformations which we want to introduce now.

3 From ordinal patterns to the ordinal transformation

Normalization and limiting. Given $i_1^\tau = i_1^\tau(t), i_2^\tau = i_2^\tau(t), i_3^\tau = i_3^\tau(t), \ldots$, for each $d \in \mathbb{N}$ formula (7) defines numbers $n_1, n_2, \ldots, n_d, \ldots$, each providing some information on the ordinal structure of the time series at time t under consideration of some finite past. The resulting numbers, however, are on different scales. In order to remove this disadvantage, we divide n_d by the natural scaling constant $(d+1)!$. This in particular allows to consider an infinite past by limiting the scaled numbers.

Theorem 5. *The assignment*

$$(i_1, i_2, i_3, \ldots) \mapsto \nu((i_1, i_2, i_3, \ldots)) = \lim_{d \to \infty} \frac{n_d}{(d+1)!} = \sum_{l=1}^{\infty} \frac{i_l}{(l+1)!}$$

with $n_d = \sum_{l=1}^{d} i_l \frac{(d+1)!}{(l+1)!}$ defines a surjection ν from the set $\mathcal{I} = \{0,1\} \times \{0,1,2\} \times \{0,1,2,3\} \times \ldots$ onto the interval $[0,1]$ having the following properties:

(i) *ν turns the lexicographic order on $\mathcal{I}$ into the usual order on $[0,1]$.*

(ii) *ν is continuous with respect to the product topology on $\mathcal{I}$ and the usual topology on $[0,1]$.*

(iii) *$\nu((i_1, i_2, i_3, \ldots)) = \nu((j_1, j_2, j_3, \ldots))$ iff $i_l = j_l$ for all $l \in \mathbb{N}$ or there exist some $k, i \in \mathbb{N}$ with $i_l = j_l$ for $l < k$ and $\{(i_k, i_{k+1}, \ldots), (j_k, j_{k+1}, \ldots)\} = \{(i,0,0,0,\ldots), (i-1, k+1, k+2, k+3, \ldots)\}$.*

(iv) For all $(i_1,\ldots,i_d)\in\mathcal{I}_d$*:* $\nu(\{(j_1,j_2,j_3,\ldots)\in\mathcal{I}\mid(j_1,\ldots,j_d)=(i_1,\ldots,i_d)\})=\left[\sum_{l=1}^{d}\frac{i_l}{(l+1)!},\left(\sum_{l=1}^{d}\frac{i_l}{(l+1)!}\right)+\frac{1}{(d+1)!}\right]$.

Proof. Indeed, the real number $\frac{n_d}{(d+1)!}$ lies in the interval $[0,1]$ and since $\frac{n_d}{(d+1)!}=\frac{1}{(d+1)!}\sum_{l=1}^{d}i_l\frac{(d+1)!}{(l+1)!}=\sum_{l=1}^{d}\frac{i_l}{(l+1)!}$ as a function in d is monotone non-decreasing, the number

$$\lim_{d\to\infty}\frac{n_d}{(d+1)!}=\sum_{l=1}^{\infty}\frac{i_l}{(l+1)!}$$

is well-defined. Moreover, the statement in (i) that the lexicographic order is turned into the usual one is an immediate consequence of Lemma 4, and the proof of (ii) is standard. As a consequence of (ii) the image of ν is closed. Since by Lemma 4 $(i_1,i_2,\ldots,i_d)\mapsto\frac{n_d}{(d+1)!}$ maps to $\{0,\frac{1}{(d+1)!},\frac{2}{(d+1)!},\ldots,\frac{(d+1)!-1}{(d+1)!}\}$, the ν-image of the set of sequences ending with $000\ldots$ is dense in $[0,1]$. Therefore, ν is surjective, and (i) shows that sequences as described in (iii) have the same image. The other direction of (iii) is an immediate consequence of (i). Now (iv) follows from $\nu((i_1,i_2,\ldots,i_d,0,0,0,\ldots))=\sum_{l=1}^{d}\frac{i_l}{(l+1)!}$ and

$$\begin{aligned}&\nu((i_1,\ldots,i_d,d+1,d+2,\ldots))-\nu((i_1,\ldots,i_d,0,0,\ldots))\\=\;&\nu((\underbrace{0,\ldots,0,0}_{d\text{ times}},d+1,d+2,\ldots))\\=\;&\nu((\underbrace{0,\ldots,0}_{d-1\text{ times}},1,0,0,\ldots))=\frac{1}{(d+1)!}\,.\end{aligned}$$

□

Definition of the ordinal transformation. Now we are able to say what we mean by the ordinal transformation.

Definition 6. *The* ordinal transformation of order $d\in\{1,2,3,\ldots,\infty\}$ *assigns to a time series* $(x_t)_{t\in\mathbb{Z}}$ *the time series* $(\nu_d^\tau(t))_{t\in\mathbb{Z}}$ *with values*

$$\nu_d^\tau(t)=\sum_{l=1}^{d}\frac{i_l^\tau(t)}{(l+1)!}\in[0,1].$$

From the viewpoint of practical data analysis one only needs the ordinal transformation for some (not too large) $d\in\mathbb{N}$. Note that usual floating point numbers allow to represent ordinal patterns (only) up to $d=30$. In the upcoming section we shall see visualizations resulting from the transformations for $d=3$ and $d=7$, respectively.

However, the ordinal transformation with $d=\infty$ defined for $T=\mathbb{Z}$ is interesting for theoretical reasons, in particular, for the 'ordinal' modelling of time series. Therefore, let us characterize the ordinal transformation for $d=\infty$ from the viewpoint of symbolic dynamics. We want to examine in particular in which way the ordinal patterns are arranged in the interval $[0,1]$.

Structure of the ordinal transformation. In fact we are interested in the structure of the map $\boldsymbol{\nu} : \mathbb{R}^\infty \longrightarrow [0,1]$ given by

$$\begin{aligned}\boldsymbol{\nu}((v_0, v_1, v_2, \ldots)) &= \sum_{l=1}^{\infty} \frac{i_l(\boldsymbol{\pi}((v_0, v_1, \ldots, v_l)))}{(l+1)!} \\ &= \sum_{l=1}^{\infty} \frac{\#\{r \in \{0, 1, \ldots, l-1\} \mid v_l \geq v_r\}}{(l+1)!}.\end{aligned}$$

and providing

$$\nu_\infty^\tau(t) = \boldsymbol{\nu}((x_t, x_{t-\tau}, x_{t-2\tau}, \ldots)).$$

Remark. Note that $i_l(\boldsymbol{\pi}((v_0, v_1, \ldots, v_k)))$ is constant for $k \geq l$.

By Theorem 5, the map $\boldsymbol{\nu}$ is surjective, i.e.

(A1) $$\boldsymbol{\nu} \text{ maps } \mathbb{R}^\infty \text{ onto } [0,1].$$

Moreover it follows from the proof of Theorem 5 that $\boldsymbol{\nu}((v_0, v_1, v_2, \ldots)) = 0$ iff $v_0 > v_1 > v_2 > \ldots$ and $\boldsymbol{\nu}((v_0, v_1, v_2, \ldots)) = 1$ iff $v_0 \leq v_1 \leq v_2 \leq \ldots$, respectively; in words: the ordinal transformation maps strictly monotone decreasing sequences to 0 and monotone increasing sequences to 1, arranging the other 'infinite' ordinal patterns in between. Similarly as mentioned at the end of Section 2 for the finite level, it appears natural to have maximum distance for these two (extreme opposite) patterns. We continue studying the arrangement of the patterns in order to extract reasonable statements for a characterization of $\boldsymbol{\nu}$.

Each ordinal pattern $\pi = (r_0, r_1, \ldots, r_d)$ of order d describes a piece of the $\mathbb{R}^\infty$ consisting of points with the same ordinal structure for the first $d+1$ components. Excluding equality of some of these components, this piece is given by

$$D_\pi = \{\mathbf{v} \in \mathbb{R}^\infty \mid \boldsymbol{\pi}^*(P_d(\mathbf{v})) = (r_0, >, r_1, >, \ldots, >, r_{d-1}, >, r_d)\},$$

where P_d denotes the projection providing the vector of the first $d+1$ components of an $\mathbf{v} \in \mathbb{R}^\infty$. Clearly, D_π is an open convex set and its closure is given by

$$\mathrm{cl}(D_\pi) = \{\mathbf{v} \in \mathbb{R}^\infty \mid \boldsymbol{\pi}^*(P_d(\mathbf{v})) = (r_0, *_1, r_1, *_2, \ldots, *_d, r_d); *_i \in \{>, =\}\}.$$

Now equality in $P_d(\mathbf{v})$ is allowed. In the following, we also want to consider the trivial permutation (0) of *order* 0 on $\{0\}$ mapping 0 to 0 with the natural assignment $D_0 = \mathbb{R}^\infty$.

Let us discuss in which way $\boldsymbol{\nu}$ preserves the geometric structure of the $\mathbb{R}^\infty$. First of all, by Theorem 5 (iv)

(8) $$\boldsymbol{\nu}(D_\pi) = \left[\sum_{l=1}^{d} \frac{i_l(\pi)}{(l+1)!}, \left(\sum_{l=1}^{d} \frac{i_l(\pi)}{(l+1)!}\right) + \frac{1}{(d+1)!}\right]$$

for π of order d, leading to the following three statements:

(A2) $\qquad \boldsymbol{\nu}(D_\pi)$ is connected for each ordinal pattern π,

(A3) $\quad \boldsymbol{\nu}(D_{\pi_1})$ and $\boldsymbol{\nu}(D_{\pi_2})$ have common lengths for π_1, π_2 of common orders,

(A4) $\quad \sharp(\boldsymbol{\nu}(D_{\pi_1}) \cap \boldsymbol{\nu}(D_{\pi_2})) \leq 1$ for different π_1, π_2 of common order.

(A1)-(A4) restrict severely the number of possible arrangements of ordinal patterns. In particular $\boldsymbol{\nu}(D_\pi) = [0,1]$ for $\pi = (0)$, $\boldsymbol{\nu}(D_\pi) \in \{[0, \frac{1}{2}], [\frac{1}{2}, 1]\}$ for $\pi \in \{(0,1),(1,0)\}$ and so on. The next statement which is a bit more complicated expresses how the ordinal transformation arranges patterns recursively.

For π_1, π_2 of order $d \geq 1$ and π of order $d-1$ with $D_{\pi_1}, D_{\pi_2} \subset D_\pi$:

(A5) $\; P_d(\mathrm{cl}(D_{\pi_1}) \cap \mathrm{cl}(D_{\pi_2}))$ has dimension $d \Longrightarrow \boldsymbol{\nu}(D_{\pi_1}) \cap \boldsymbol{\nu}(D_{\pi_2}) \neq \emptyset$.

This means that $\boldsymbol{\nu}$ preserves non-disjointness of two pieces of order d inside a piece of order $d-1$ if the intersection is 'as large as possible'. Indeed, under the above assumption the dimension of $P_d(\mathrm{cl}(D_{\pi_1}))$ and $P_d(\mathrm{cl}(D_{\pi_2}))$ is $d+1$; and since for $\pi_1 \neq \pi_2$ the sets D_{π_1} and D_{π_2} are disjoint, $P_d(\mathrm{cl}(D_{\pi_1}) \cap \mathrm{cl}(D_{\pi_2})) = P_d(\mathrm{cl}(D_{\pi_1})) \cap P_d(\mathrm{cl}(D_{\pi_2}))$ has dimension not greater than d. It holds $i_l(\pi_1) = i_l(\pi_2)$ for $l = 1, 2, \ldots, d-1$, and if the dimension considered is d, there are only two indices $i, j \leq d$, where between D_{π_1} and D_{π_2} the relation $v_i > v_j$ flips to $v_i < v_j$. Clearly, $i = d$ or $j = d$, hence $|i_d(\pi_1) - i_d(\pi_2)| = 1$, implying $\boldsymbol{\nu}(D_{\pi_1}) \cap \boldsymbol{\nu}(D_{\pi_2}) \neq \emptyset$ (see Theorem 5 (iv)).

Finally, according to the proof of Theorem 5 the following is valid:

For π of order $d \geq 0$ and $\mathbf{v} = (v_i)_{i=1}^{\infty} \in D_\pi$:

(A6) $\quad v_0, v_1, \ldots, v_d > v_{d+1} > v_{d+2} > v_{d+3} > \ldots \Longrightarrow \boldsymbol{\nu}(\mathbf{v}) = \min \boldsymbol{\nu}(D_\pi)$.

Characterization of the ordinal transformation. For a map $\boldsymbol{\nu}$ from the $\mathbb{R}^\infty$ into $[0,1]$ identifying equal ordinal structure but discriminating different one, and preserving as much geometrical structure as possible, the conditions (A1) - (A4) are reasonable, and (A5) still appears naturally. (A6), however, is the only purely arbitrary setting. We motivate the choice of (A6) as follows: For a time series $(x_t)_{t \in \mathbb{Z}}$ and some delay τ, a sequence $(x_{t-l\tau})_{l=0}^{\infty}$ is monotone decreasing iff for the corresponding time series part time order and order of values coincide, which we consider as the 'normal' case. The ordinal transformation yields in somehow a 'distance' from this 'normal' behavior. More precisely, (A6) says that inside D_π sequences $\mathbf{v}$ with the smallest number of inversions provide the smallest $\boldsymbol{\nu}(\mathbf{v})$.

The following theorem characterizes the ordinal transformation by (A1)-(A6).

Theorem 7. *Let $\boldsymbol{\nu} : \mathbb{R}^\infty \to [0,1]$ be a map satisfying (A1)-(A6), then*

$$\boldsymbol{\nu}(\mathbf{v}) = \sum_{l=1}^{\infty} \frac{\#\{r \in \{0, 1, \ldots, l-1\} \mid v_l \geq v_r\}}{(l+1)!}$$

for all $\mathbf{v} = (v_0, v_1, v_2, \ldots) \in \mathbb{R}^\infty$ with mutually different components.

Proof. Let $\boldsymbol{\nu}$ be given with (A1)-(A6). Using induction, we first verify (8). Assume that (8) is shown for all π of some order $d = k$. Then choose such a $\pi = (r_0, r_1, \ldots, r_k)$. The length of the interval $\boldsymbol{\nu}(D_\pi)$ is $\frac{1}{(k+1)!}$ and its smallest element $\sum_{l=1}^{k} \frac{i_l(\pi)}{(l+1)!}$. Given $i \in \{0, 1, 2, \ldots, k+1\}$, let σ_i be the permutation of order $k+1$ defined by $i_l(\sigma_i) = i_l(\pi)$ for $l = 1, 2, \ldots, k$ and $i_{k+1}(\sigma_i) = i$. Clearly, all $\boldsymbol{\nu}(D_{\sigma_i})$ are contained in $\boldsymbol{\nu}(D_\pi)$ and by (A6) their smallest elements are $(\sum_{l=1}^{k} \frac{i_l(\pi)}{(l+1)!}) + \frac{i}{(k+2)!}$. Moreover, the set $P_{k+1}(\mathrm{cl}(D_{\sigma_i}) \cap \mathrm{cl}(D_{\sigma_{i+1}})) = \{(v_0, v_1, \ldots, v_{k+1}) | v_{r_1} > \ldots > v_{r_{k-i}} = v_{k+1} > v_{r_{k-i+1}} > \ldots > v_{r_k}\}$ has dimension $k+1$ for $i = 0, 1, \ldots, k$, thus by (A2)-(A5) $\boldsymbol{\nu}(D_{\pi_i}) = [(\sum_{l=1}^{k} \frac{i_l(\pi)}{(l+1)!}) + \frac{i}{(k+2)!}, (\sum_{l=1}^{k} \frac{i_l(\pi)}{(l+1)!}) + \frac{i+1}{(k+2)!}]$ for all $i = 0, 1, \ldots, k+1$. Now the validity of (8) is obvious for $d = k+1$.

If $\mathbf{v}$ has mutually different components, then there exists a sequence $\pi_1, \pi_2, \pi_3, \ldots$ of increasing order with $\mathbf{v}$ in all D_{π_i}. Since the length of $\boldsymbol{\nu}(D_{\pi_i})$ converges to 0, the point $\boldsymbol{\nu}(\mathbf{v})$ is uniquely determined by knowing all $\boldsymbol{\nu}(D_\pi)$. $\square$

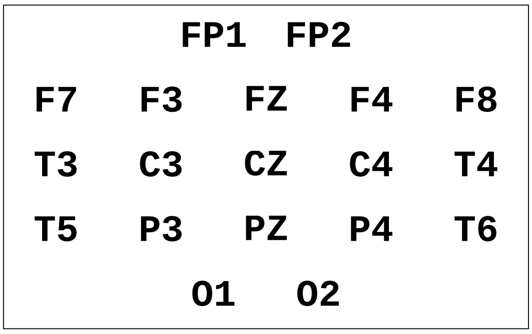

Figure 2: 10-20-system

4 Illustration

In this section we want to illustrate ordinal time series analysis by looking at some EEG data. Note that the above methods are particularly provided for the exploration of long time series reflecting qualitatively different states of complex systems. Such time series are indeed given by the electroencephalogram (EEG) measuring the electrical activity of the brain by the use of electrodes.

Figure 3 visualizes six different EEG samples which have been recorded from scalp electrodes with a sampling rate of $256Hz$. The electrode placement (according to the international 10-20-system) is sketched in Figure 2: FP1, FP2 are in the front of the head, T3 and T4 on the left and the right side, respectively. Note that the six samples have been collected from the same person at different times (EEG 1: recorded from a boy at an age of 8 years, EEG 2: recorded three years later) and from different EEG channels, respectively, as indicated in Figure 3. All samples are 108 s = 27648 data points long. The plots in the top row show the original data sets and 3 s long parts of them.

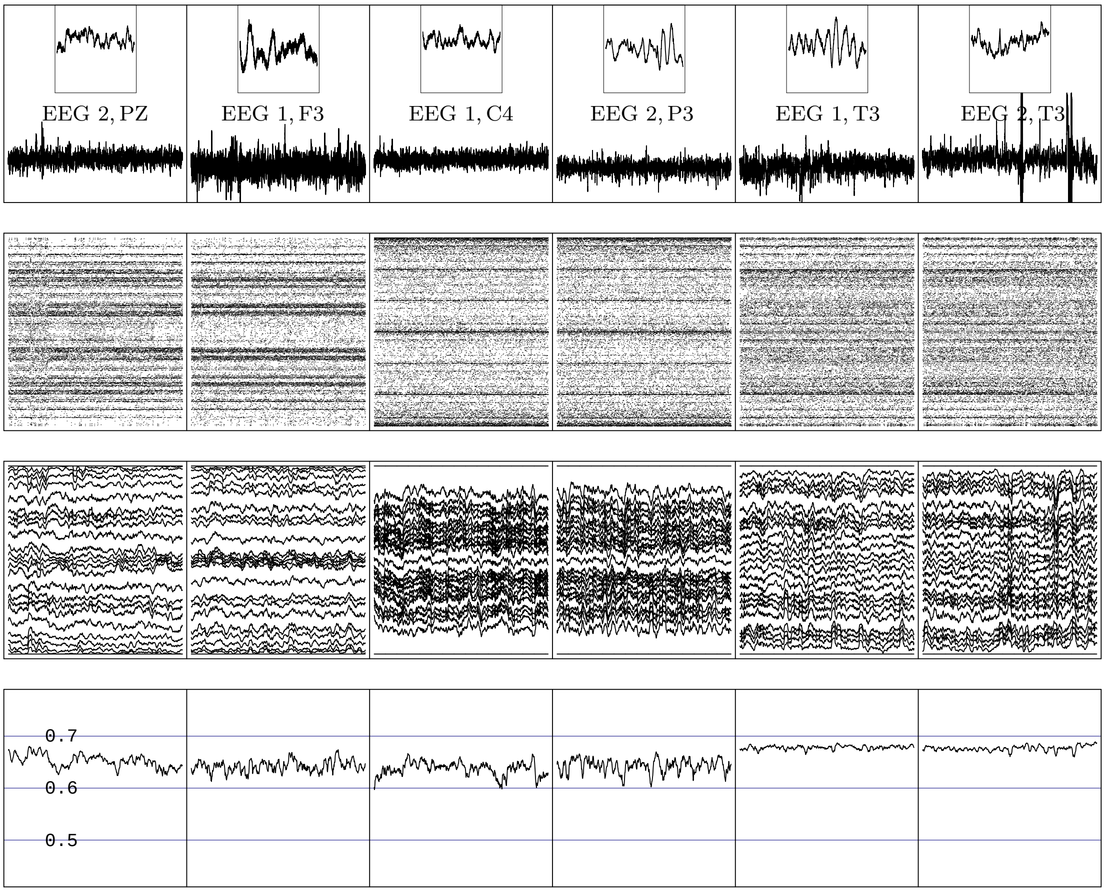

Figure 3: EEG examples

In the second row the time series obtained from the original time series by the ordinal transformation of order $d = 7$ are provided. Here and in the calculations behind the graphics in the third and fourth row the delay τ is equal to 1. The extremely truncated representation in time direction allows to make out black and white horizontal 'stripes' suggesting stationarity of the underlaying 'ordinal process'. Looking at the transformed data one finds strong resemblances between the first and the second plot, the third and the fourth plot and the fifth and the sixth plot, respectively, while differences between these three pairs appear very clearly. This only partially matches with observations from the original data: the signals belonging to the third and fourth plot particularly show some longer and steeper wave forms whereas the first two signals are more 'noisy'. However, only by looking at the original data one probably would not have separated the fifth and the sixth part from the third and the fourth part. Note that usually contrasts between qualitatively different EEG signals are strengthened by the ordinal transformation (see [5]).

Clearly, the different impressions from the plots of the second row are related to differences between the pattern distributions. This is what the plots in the third row from above reflect: for sliding time windows of 2 seconds (i.e. 512 data points) they show the time-dependent pattern distributions for order $d = 3$. The spaces between succeeding curves represent the relative frequency of ordinal patterns, with the pattern $n_3 = 0$ associated to the bottom space, $n_3 = 1$ to the space between the first and the second curve from below, ..., and $n_3 = 23$ to the top space. The observations concerning similarities and dissimilarities between different parts we made above still hold for these representations, however the distributions allow a quantification. Note moreover that pattern distributions directly allow to make qualitative statements about the original signal: for example a vaster occurrence of the 'bottom' and 'top' patterns $n_3 = 0$ and $n_3 = 23$, respectively, indicates a vaster occurrence of strictly monotone increasing and monotone decreasing parts, respectively, within the original signal.

The plots in the bottom row of Figure 3 finally provide the (empirical) permutation entropies on the base of the distributions represented in the third row. (The scale drawn is normalized to have maximal possible value 1.) Note that by definition the permutation entropy for some d is no more than the Shannon entropy of the corresponding ordinal pattern distribution (see [1]). As already mentioned it has been taken as a measure for the complexity of a system behind a time series, leading to good results in the detection of epileptic activity in EEG signals (compare [4]). In our example the permutation entropy is on a nearly identic level, in particular for the first four data sets (showing more fluctuations than the two other ones). Clearly, the permutation entropy does not reflect all features of the ordinal pattern distribution.

That's why we propose to consider not only the permutation entropy but the whole ordinal structure of a time series. As illustrated above, the use of the ordinal transformation allows to separate qualitatively different 'states' within the original signal(s). Taking distributions of patterns into account, one may quantify similarity and dissimilarity, respectively, in order to separate these states automatically. For a first step into this direction see Keller and Wittfeld [6]. Having in mind that the data sets considered above have been collected from the same person but at different times and from different EEG channels, the following question arises: Does there exist a basic (individual) repertoire of 'ordinal' states of brain activity?

5 Invariance properties

Characteristics on the base of ordinal pattern distributions like the permutation entropy are interesting in particular by two reasons: On the one hand, ordinal patterns are relatively robust with respect to small noise, affecting only few of the order relations between values, and on the other hand they clearly do not change under monotone transformations which means that they are invariant in particular with respect to different scalings or offsets of the original signal. Here we want to have a closer look at the latter fact stating invariance properties of ordinal symbolic dynamics. This will also lead us to a special characterization of the decomposition $\mathfrak{D}_d^{ord}$.

We say that a partition $\mathfrak{D}$ of the $\mathbb{R}^{d+1}$ is *invariant* under a map f on $\mathbb{R}$ if the following holds: $\mathfrak{D}$ separates the vectors $(v_0, v_1, \ldots, v_d)$ and $(w_0, w_1, \ldots, w_d)$ iff $\mathfrak{D}$ separates $(f(v_0), f(v_1), \ldots, f(v_d))$ and $(f(w_0), f(w_1), \ldots, f(w_d))$. It is obvious that

$$\mathfrak{D}_d^{ord} \text{ is invariant under strictly monotone maps.}$$

In order to characterize $\mathfrak{D}_d^{ord}$ by other invariance properties, we first show the following statement:

Lemma 8. *Each finite partition of* $\mathbb{R}$ *which is invariant under translations is trivial.*

Proof. Assume the existence of a translation-invariant partition $\mathfrak{E} = \{E_0, E_1, \ldots, E_p\}$ of $\mathbb{R}$ for $p \geq 1$. Further, assume that $0 \in E_0$. We first note that

$$E_0 \text{ is a subgroup of } (\mathbb{R}, 0, +). \tag{9}$$

Let $v, w \in E_0$ be different. Then $0 = v + (-v)$ and $-v = 0 + (-v)$ are not separated by $\mathfrak{E}$, showing that $-v \in E_0$. On the other hand, 0 and v are not separated by $\mathfrak{E}$, hence $w = 0 + w$ and $v + w$ are not, implying $v + w \in E_0$.

Let $v \in E_q$ for $q \neq 0$ and $w \in \mathbb{R}$. Then $0, w$ are not separated by $\mathfrak{E}$ iff $v = 0 + v$ and $w + v$ are not separated by $\mathfrak{E}$. Therefore,

$$E_q = v + E_0 \text{ for all } v \in E_q \text{ and all } q = 0, 1, \ldots, p. \tag{10}$$

An immediate consequence of (10) is that each E_q must be infinite and unbounded and that

$$v_1 - v_2 \in E_0 \text{ for all } v_1, v_2 \in E_q \text{ and all } q = 0, 1, \ldots, p. \tag{11}$$

Now fix some $u \in \mathbb{R} \setminus E_0$. Further, for $q = 1, 2, \ldots, p+2$ let $u_q = \frac{u}{q \prod_{j=1}^{p+1} j}$. We show that the u_q must be in mutually different pieces of $\mathfrak{E}$, which clearly is impossible. If u_{q_1} and u_{q_2} were not separated by $\mathfrak{E}$ for $q_1 < q_2$, then by (11) it would hold $E_0 \ni u_{q_1} - u_{q_2} = u \left(\frac{1}{q_1 \prod_{j=1}^{p+1} j} - \frac{1}{q_2 \prod_{j=1}^{p+1} j} \right) = u \left(\frac{q_2 - q_1}{q_1 q_2 \prod_{j=1}^{p+1} j} \right)$. Since $q_2 - q_1$ divides $\prod_{j=1}^{p+1} j$, the number u would be a multiple of $u_{q_1} - u_{q_2}$, hence by (9) it would lie in E_0, contradicting our assumption. $\square$

By the fact that $x \mapsto \ln x$ forms a group isomorphism from $(\mathbb{R}, 0, +)$ onto $(\mathbb{R}^+, 1, \cdot)$ with $\mathbb{R}^+ =]0, \infty[$ one immediately obtains

Corollary 9. *Each finite partition of* $\mathbb{R}^+$ *which is scaling-invariant, i.e. invariant under multiplication with any positive number, is trivial.*

In the following a partition $\mathfrak{D}$ is said to be *coarser* than a partition $\mathfrak{E}$ if each $E \in \mathfrak{E}$ is contained in some $D \in \mathfrak{D}$. This includes equality of the partitions. First we want to give a characterization of the partition $\mathfrak{D}_1^{ord}$ of $\mathbb{R}^2$ consisting of the diagonal $\{(x, x) | x \in \mathbb{R}\}$ and the two parts of $\mathbb{R}^2$ separated by the diagonal.

Corollary 10. *A finite partition of the $\mathbb{R}^2$ is coarser than $\mathfrak{D}_1^{ord}$ iff it is translation- and scaling-invariant.*

Proof. The one direction is obvious. In order to verify the other direction, let $\mathfrak{D}$ be translation- and scaling-invariant, and let (v_0, v_1) and (w_0, w_1) be two vectors not separated by $\mathfrak{D}_1^{ord}$.

If $v_0 = v_1$, hence $w_0 = w_1$, then (v_0, v_1) and (w_0, w_1) lie on the diagonal of $\mathbb{R}^2$, which is ('translation')-invariant under addition of vectors (u, u) for $u \in \mathbb{R}$. The application of Lemma 8 to the restriction of $\mathfrak{D}$ to the diagonal shows that (v_0, v_1) and (w_0, w_1) are not separated by $\mathfrak{D}$.

If $v_0 < v_1$, hence $w_0 < w_1$ (or $v_0 > v_1$, hence $w_0 > w_1$), the line through (v_0, v_1) parallel to the diagonal and the ray $\{\lambda(w_0, w_1) \mid \lambda \in \mathbb{R}^+\}$ have a common point. Since $\mathfrak{D}$ does not separate points on the diagonal and on the ray by Lemma 8 and Corollary 9, respectively, (v_0, v_1) and (w_0, w_1) are not separated by $\mathfrak{D}$. □

A statement similar to Corollary 10 for $d > 1$ fails to be valid. For example, let $d = 2$, let $D = \{(x_0, x_1, x_2) \in \mathbb{R}^3 \mid x_0 - x_1 > x_1 - x_2 > 0\}$ and $\mathfrak{D} = \{D, \mathbb{R}^3 - D\}$. Then e.g. $(3, 1, 0)$ and $(3, 2, 0)$ are separated by $\mathfrak{D}$, but not by $\mathfrak{D}_2^{ord}$, and $\mathfrak{D}$ is translation- and scaling-invariant. However, strengthening the scaling-invariance, by Theorem 11 we have an adequate statement for general d.

Before Definition 3 we spoke about ordinal symbolic dynamics as symbolic dynamics with finite partitions of the $\mathbb{R}^{d+1}$ only depending on the ordinal structure of a vector. Such partitions are exactly described by each of the following equivalent conditions (i), (ii), (iii).

Theorem 11. *Let $\mathfrak{D}$ be a finite partition of the $\mathbb{R}^{d+1}$ for $d \geq 1$. Then the following statements are equivalent:*

(i) $\mathfrak{D}$ is coarser than $\mathfrak{D}_d^{ord}$.

(ii) $\mathfrak{D}$ is invariant under strictly increasing maps.

(iii) $\mathfrak{D}$ is invariant under translations and strictly increasing piecewise linear continuous maps with at most two pieces.

Proof. The implications (i) $\Longrightarrow$ (ii) and (ii) $\Longrightarrow$ (iii) are obvious. In order to verify (iii) $\Longrightarrow$ (i), suppose that (iii) is valid. The proof uses a special type of strictly increasing piecewise linear continuous maps with at most two pieces defined by

$$L^{z,\lambda}(x) = \begin{cases} x & \text{for } x \leq z \\ z + \lambda(x - z) & \text{else} \end{cases}$$

for $z \in \mathbb{R}$ and $\lambda \in \mathbb{R}^+$, and providing a map $\mathbf{L}^{z,\lambda} : \mathbb{R}^{d+1} \hookleftarrow$ with $\mathbf{L}^{z,\lambda}(\mathbf{x}) = \mathbf{L}^{z,\lambda}((x_i)_{i=0}^d) = (L^{z,\lambda}(x_i))_{i=0}^d$. Clearly, $\mathbf{L}^{z,1}(\mathbf{x}) = \mathbf{x}$ for all $z \in \mathbb{R}$, and $\mathbf{L}^{z,\lambda}(\mathbf{x}) = \mathbf{x}$ if $x_i \leq z$ for all $i = 0, 1, \ldots, d$. When z and $\mathbf{x} = (x_i)_{i=0}^d \in \mathbb{R}^{d+1}$ are given such that

$x_i > z$ for at least one i, the family $\{\mathbf{L}^{z,\lambda} \mid \lambda \in \mathbb{R}^+\}$ acts on the ray $\{\mathbf{L}^{z,\lambda}(\mathbf{x}) \mid \lambda \in \mathbb{R}^+\}$ like the multiplication with λ on $\mathbb{R}^+$.

Now let $\mathbf{v} = (v_0, v_1, \ldots, v_d)$ and $\mathbf{w} = (w_0, w_1, \ldots, w_d)$ be not separated by $\mathfrak{D}_d^{ord}$, i.e. let $\boldsymbol{\pi}^*(\mathbf{v}) = \boldsymbol{\pi}^*(\mathbf{w}) = (r_0, *_1, r_1, *_2, \ldots, *_{d-1}, r_{d-1}, *_d, r_d)$ be defined as above. First fix some $u \in \mathbb{R}$ such that the coordinate with index r_d of $\widetilde{\mathbf{v}} = (v_0 + u, v_1 + u, \ldots, v_d + u)$ and w_{r_d} coincide. Since $\mathfrak{D}$ is invariant under translations, one can apply Lemma 8 to see that $\mathbf{v}$ and $\widetilde{\mathbf{v}}$ are not separated by $\mathfrak{D}$ in a similar way as in the proof of Corollary 10. Using Corollary 9 analogously, we are done by providing maps $\mathbf{L}_0, \ldots, \mathbf{L}_{d-1} \in \mathcal{L} := \{\mathbf{L}^{z,\lambda} \mid z \in \mathbb{R}, \lambda \in \mathbb{R}^+\}$ with $\mathbf{w} = \mathbf{L}_0 \circ \ldots \circ \mathbf{L}_{d-1}(\widetilde{\mathbf{v}})$.

Indeed, assume that for some $k \in \{0, \ldots, d-1\}$ maps $\mathbf{L}_{k+1}, \mathbf{L}_{k+2}, \ldots, \mathbf{L}_{d-1} \in \mathcal{L}$ are defined such that w_{r_l} is equal to the coordinate with index r_l of $\mathbf{L}_{k+1} \circ \mathbf{L}_{k+2} \circ \ldots \circ \mathbf{L}_{d-1}(\widetilde{\mathbf{v}})$ for all $l = k, k+1, \ldots, d-1$. Then if $*_k$ is equal to $=$, let $\mathbf{L}_k$ be the identity. Otherwise let $\mathbf{L}_k = \mathbf{L}^{z,\lambda}$ be given with $z = w_{r_{k+1}}$ and $\lambda = \frac{w_{r_k} - w_{r_{k+1}}}{w - w_{r_{k+1}}}$, where w denotes the coordinate with index r_k of $\mathbf{L}_{k+1} \circ \mathbf{L}_{k+2} \circ \ldots \circ \mathbf{L}_{d-1}(\widetilde{\mathbf{v}})$. Then w_{r_l} coincides with the coordinate with index r_l of $\mathbf{L}_k \circ \mathbf{L}_{k+1} \circ \ldots \circ \mathbf{L}_{d-1}(\widetilde{\mathbf{v}})$ for all $l = k-1, k, \ldots, d-1$. □

Acknowledgement. We would like to thank Heinz Lauffer from the Department of Pediatric Medicine of the University Greifswald for placing the example data sets discussed in Section 4 at our disposal.

References

[1] C. Bandt, B. Pompe, Permutation entropy: A natural complexity measure for time series, *Phys. Rev. Lett.* 88 (2002), 174102.

[2] C. Bandt, G. Keller, B. Pompe, Entropy of interval maps via permutations, *Nonlinearity* 15 (2002), 1595 – 1602.

[3] P. Diaconis, *Group representations in probability and statistics*, IMS, Hayward, California, 1988.

[4] K. Keller, H. Lauffer, Symbolic analysis of high-dimensional time series, *Int. J. Bifurcation Chaos* 13 (2003), 2657 – 2668.

[5] K. Keller, M. Sinn, Ordinal analysis of time series, Preprint, Lübeck, 2005.

[6] K. Keller and K. Wittfeld, Distances of time series components by means of symbolic dynamics, *Int. J. Bifurcation Chaos* 14 (2004), 693–704.

[7] D. Knuth, *The art of computer programming. Volume 3: Sorting and searching*, Addison Wesley, Reading, Massachusetts, 1973.