

Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Operator-theoretic identification of closed
sub-systems of dynamical systems

by

Oliver Pfante and Nihat Ay

Preprint no.: 4

2015



Operator-theoretic identification of closed sub-systems of dynamical systems

Oliver Pfante ^{*1} and Nihat Ay ^{†1,2,3}

¹Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany

²Department of Mathematics and Computer Science, Leipzig University, PF 100920, 04009 Leipzig, Germany

³Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA

Abstract

A central problem of dynamical systems theory is to identify a reduced description of the dynamical process one can deal easier. In this paper we present a systematic method of identifying those closed sub-systems of a given discrete time dynamical system in the frame of operator theory. It is shown that this problem is closely related to finding invariant sigma algebras of the dynamics.

Index Terms – closed sub-dynamics, sigma algebras, operator theory.

1 Introduction

Classical scientific models often derive their success from providing a closed description of some section of reality. The Schrödinger equation, for instance, describes the quantum mechanical behavior of the hydrogen atom without being concerned with the fact that the proton constituting its nucleus is not elementary, but is composed of quarks. However, the mechanism of quark confinement makes a consistent description at the atomic level possible. Likewise, Newtonian dynamics models the sun, the planets and their moons as point masses and can then capture their dynamics without having to consider their rather complicated internal structure in terms of solid state physics or fluid or gas dynamics. In biology, the Hodgkin-Huxley equations and their variants operate successfully at the cellular level without taking the molecular details of the ion channels into account. The Wright-Fisher and other models of population genetics express genetic drift in large populations in terms of Fokker-Planck type equations without the complicated details of the mechanisms of genetic inheritance at the DNA level.

However, such a reducing scheme is not always easily available. Neuroscience or cell biology provide examples where it may not even be clear what the appropriate levels are. And even in physics, the transition from the atomic or molecular scale to a hydrodynamic description in the case of non-Newtonian fluids or phenomenological descriptions in materials science provide some of the most difficult problems that science faces.

This raises the question how we can identify such a level given some microscopic description of a system. We refer to a level as a new macroscopic dynamics derived from the microscopic one by means of a coarse-graining or aggregation of micro-states. More precisely, consider a dynamical system $\phi : X \rightarrow X$ on a state space X where ϕ can be a measurable map, or, more generally, a Markov kernel. Suppose we have an

*pfante@mis.mpg.de

†nay@mis.mpg.de

operator $\pi : X \rightarrow \hat{X}$ – for instance a coarse-graining, aggregation, averaging, etc. – of the lower, microscopic level X onto an upper level \hat{X} . As the time-discrete dynamics evolves on the lower level by means of an iteration of the map ϕ , an induced dynamics can be observed on the upper state space \hat{X} . We say that the upper level is *closed* if it can be also represented by a dynamical system: there is a measurable map or a Markov kernel, etc. $\hat{\phi} : \hat{X} \rightarrow \hat{X}$ such that $\pi \circ \phi = \hat{\phi} \circ \pi$.

$$\begin{array}{ccc} X & \xrightarrow{\phi} & X \\ \downarrow \pi & & \downarrow \pi \\ \hat{X} & \xrightarrow{\hat{\phi}} & \hat{X} \end{array}$$

Figure 1: Basic setup of multilevel dynamical system.

The problem of level identification is extensively studied for several time discrete dynamical systems on finite state spaces. N. Israeli and N. Goldenfeld [12] studied this problem for cellular automata. The authors make use of local aggregations of nearest neighbors of the states of one-dimensional cellular automata to obtain computational reduced descriptions of all cellular automata in Wolfram’s classification scheme. A vast amount of literature is available where the operator ϕ in Fig. (1) is a time-discrete, stationary Markov process on a finite state space X . The problem of a level identification turns into the one of finding lumpings $\pi : X \rightarrow \hat{X}$ of the states in X such that the macroscopic dynamics $\hat{\phi} : \hat{X} \rightarrow \hat{X}$ turns out to be Markovian again. [17] proves that the micro-process $\phi : X \rightarrow X$ must be invariant with respect to the action of certain permutations of the states in X in order to obtain a Markovian macro-dynamics $\hat{\phi} : \hat{X} \rightarrow \hat{X}$ as well. In [5] this method of finding possible lumping was applied to accelerate convergence of Markov Chain Monte Carlo techniques. [6] deals with the problem of finding aggregations for ϕ representing a nearly uncoupled Markov chain on a finite state space X . A systematic algorithm for finding all possible lumpings of a stationary Markov process $\phi : X \rightarrow X$ on a finite state space X such that the resulting process $\hat{\phi} : \hat{X} \rightarrow \hat{X}$ is Markovian again and the diagram in Fig. (1) commutes was developed by Görnerup and Jacobi in [10] and [9] whose method makes use of certain level sets of the right eigenvectors of the stochastic matrix associated to ϕ . In [16] we introduced an information-theoretic measure not only to find aggregations π whose corresponding macro-dynamics $\hat{\phi}$ is Markovian but also to quantify to what extent the macro-level deviates from being Markovian otherwise. Beside these works, which take only care of possible strong lumpings of the Markov process ϕ – i.e., lumpings such that the macro process $\hat{\phi}$ is Markovian again no matter what choice is made for the initial distribution on X – there are some interesting theoretical considerations, elaborated in the book of Kemeny and Snell [13] whose outline is based on [3], when one requires only that at least for one initial distribution the aggregation leads to a Markov chain.

All these approaches work only for finite state spaces X where the stationary Markov process $\phi : X \rightarrow X$ can be represented by a stochastic matrix. The aggregations methods do not apply or even still make any sense if one allows for infinite state spaces X . This paper is entirely devoted to tackle this problem: finding possible aggregations $\pi : X \rightarrow \hat{X}$ for the dynamical systems $\phi : X \rightarrow X$ on an infinite state space X which yield a self-contained macro-process $\hat{\phi} : \hat{X} \rightarrow \hat{X}$ in Fig. (1).

Operator theory turns out to be an ideal conceptual frame not only providing the tools which lead to a solution of the sketched problem but also providing a guideline for the choice of the notions itself: What kind of infinite spaces X should be considered? Which time-discrete dynamical systems $\phi : X \rightarrow X$ shall be considered? Which requirements should the aggregation π fulfill? In the infinite setting there are no canonical answers to these questions and choices need to be made.

In the present paper, we proceed as follows. In the second section we introduce the operator theoretical framework in which our approach is embedded: we define operators on Banach spaces, introduce σ -finite measure spaces, the Radon-Nikodym derivative of a measure with respect to a sub- σ -algebra, etc. More

important, this section contains the main result of the paper – presenting a criterion an aggregation needs to fulfill in order to induce a new, closed level in the sense of Fig. (1).

In the third section we apply the general results of the previous one to the special case of finite state spaces. We can reproduce all well known results on the strong lumpability of Markov chains in particular, and time-discrete, stationary dynamical systems in general, by reproving the results of [11] which are the most general ones among those which were presented above.

The fourth section takes care of the special case when the linear operator T is induced by a point transformation $\phi : X \rightarrow X$. We prove that a sub- σ -algebra \mathcal{C} induces an aggregation if and only if \mathcal{C} is invariant under the preimage of ϕ , i.e., $\phi^{-1}(\mathcal{C}) \subset \mathcal{C}$. Furthermore, we develop a general procedure to obtain such invariant σ -algebras by finding eigenfunctions of the linear operator T . Every σ -algebra obtained from this eigenfunction method provides a factor, but the opposite way does not hold true – there are factors which are not induced from a family of eigenfunctions. A counterexample is obtained from ergodicity theory which also gives us an example of an invertible operator T , induced by a point transformation, whose corresponding macro-operator \hat{T} in Fig. (1) is no longer invertible at all. This phenomenon has its roots in our infinite setting, because for finite state spaces a surjective aggregation of a bijective linear process leads to a bijective macro-process as well, see for instance [16].

The last section deals with Markov kernels. We consider those which are uniquely ergodic, a notion we shall introduce. Those Markov kernels have a family of eigenvectors, called root system, whose set Γ of eigenvalues is a subgroup of the unit circle S^1 . Similar to Galois extensions in group theory we construct a bijection between subgroups of Γ and sub- σ -algebras which provide a factor of the dynamical system.

2 General Theory

This section does not only provide a rather general solution to the problem of level identification in time-discrete dynamical systems in general but also the conceptual frame in which this work is done. We start with a precise definition of the state spaces X we consider exclusively in the present work.

Definition 2.1. We call a measure space X with σ -algebra \mathcal{B} and measure μ to be σ -finite if there is a countable set $\{U_n : U_n \in \mathcal{B} \text{ and } \mu(U_n) < \infty \text{ for all } n \in \mathbb{N}\}$ such that $U_n \subset U_{n+1}$ and $X = \bigcup_{n \in \mathbb{N}} U_n$.

Remark 2.1. In the sequel we assume that the measure space (X, \mathcal{B}, μ) is complete, i.e., if $S \subset N \in \mathcal{B}$ with $\mu(N) = 0$, then also $S \in \mathcal{B}$.

Instead of a time-discrete dynamics $\phi : X \rightarrow X$, we want to consider those dynamical systems which are defined on the L^p space $L^p(\mathcal{B}, \mu)$, that is, on the Banach space of all measurable functions whose p -th power is integrable. We substitute ϕ by a linear operator $T : L^p(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu)$. Instances where this setting applies are point transformations $\phi : X \rightarrow X$ which induce on $L^p(\mathcal{B}, \mu)$ the pullback $f \mapsto f \circ \phi$ and the pullback induced by a Markov kernel. We study both examples in this paper. L^p -spaces are special instances of Banach spaces.

Definition 2.2. Let be B_1 and B_2 Banach spaces. By an operator from B_1 to B_2 we shall mean a linear mapping T whose domain $\mathcal{D}(T)$ is a subspace of B_1 and whose range $\mathcal{R}(T)$ lies in B_2 . We call an operator *bounded* if $\mathcal{D}(T) = B_1$ and the supremum

$$\sup_{\|x\|=1} \|Tx\| \tag{2.1}$$

is bounded. In this case the supremum Eq. (2.1) is called the norm of T and is denoted by $\|T\|$.

The *support* of an operator T is the subspace $\{x \in \mathcal{D}(T) : x \neq 0 \text{ then } Tx \neq 0\}$, i.e., T is injective on its support.

Absolutely not obvious is the derivation of a fruitful lumping concept in the infinite setting. It turns out that sub- σ -algebras of \mathcal{B} provide a convenient reduction scheme. The idea behind the use of sub- σ -algebras works as follows. Suppose $x \in X$. The σ -algebra \mathcal{B} is *resolved* by the information given by x . This means we can decide for every element $A \in \mathcal{B}$, which is a subset of X , whether $x \in A$ or not. If the σ -algebra \mathcal{B} is

fine enough, i.e., contains enough subsets of X such that for two points $x, y \in X$ we find two sets $A, B \in \mathcal{B}$ fulfilling $A \cap B = \emptyset$, $x \in A$ and $y \in B$ (an instance is the Borel σ -algebra on \mathbb{R}), then one can go along the opposite way: $x \in X$ is uniquely determined if we know how x resolves the σ -algebra \mathcal{B} . Lumping together two states, i.e., two different points $x, y \in X$, implies that they cannot be distinguished in the macro-space \hat{X} any longer. This lack or, respectively, reduction of information can be achieved by considering only a sub- σ -algebra \mathcal{C} of \mathcal{B} , whose resolution does not allow to discriminate between these two points x and y , that is, we have $x \in C$ if and only if $y \in C$ for all $C \in \mathcal{C}$. To link this concept with the dynamical aspects, we assume for simplicity that the operator T is induced by a point transformation $\phi : X \rightarrow X$. If the reduction, which is derived from a sub- σ -algebra \mathcal{C} , results in the one of the whole dynamics $\phi : X \rightarrow X$, it has to be respected by the map ϕ , i.e., if two points cannot be discriminated by \mathcal{C} this should also hold true for the images $\phi(x)$ and $\phi(y)$. This condition turns out to be equivalent to the one that ϕ is \mathcal{C} -measurable, that is, for all $C \in \mathcal{C}$ the preimage $\phi^{-1}(C)$ is in \mathcal{C} as well. For the general case, when the linear operator T needs not be induced by a point transformation, this insight leads to the definition of a *factor*. A well known concept in Ergodic theory already thoroughly discussed in Furstenberg's work [8] and also with a nice exposure in [18]. This is nothing else than a formalization of the presented intuitive idea of an information reduction forced by a choice of a sub- σ -algebra \mathcal{C} , which needs to be respected by the map T .

Definition 2.3. Suppose (X, \mathcal{B}, μ) and $(\hat{X}, \hat{\mathcal{B}}, \hat{\mu})$ are σ -finite measure spaces. Let be $p \in [1, \infty]$. Let be T and \hat{T} operators with domain and range in $L^p(\mathcal{B}, \mu)$ and $L^p(\hat{\mathcal{B}}, \hat{\mu})$, respectively. We call \hat{T} a *factor* of T if there exists a bounded linear operator $\Pi : L^p(\mathcal{B}, \mu) \rightarrow L^p(\hat{\mathcal{B}}, \hat{\mu})$ such that the following diagram commutes:

$$\begin{array}{ccc} L^p(\mathcal{B}, \mu) & \xrightarrow{T} & L^p(\mathcal{B}, \mu) \\ \Pi \downarrow & & \downarrow \Pi \\ L^p(\hat{\mathcal{B}}, \hat{\mu}) & \xrightarrow{\hat{T}} & L^p(\hat{\mathcal{B}}, \hat{\mu}) \end{array} \quad (2.2)$$

Finding new, closed levels in the sense of Fig. (1) is then equivalent to the problem how one can find proper factors of the dynamical system $T : L^p(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu)$. Theorem 2.2 provides a guideline by linking this problem to an invariance condition the sub- σ -algebra \mathcal{C} has to fulfill with respect to a certain projection operator on $L^p(\mathcal{B}, \mu)$.

Let $\mathcal{C} \subset \mathcal{B}$ a sub- σ -algebra of \mathcal{B} and $f \in L^1(\mathcal{B}, \mu)$. Since the measure space (X, \mathcal{B}, μ) is σ -finite, there is a unique \mathcal{C} -measurable and integrable function F , the Radon-Nikodym derivative of f , such that

$$\int_H f d\mu = \int_H F d\mu \quad \text{for all } H \in \mathcal{C}.$$

We define

$$P_{\mathcal{C}}f = F \quad \text{for all } f \in L^1(\mathcal{B}, \mu).$$

The operator $P_{\mathcal{C}} : L^1(\mathcal{B}, \mu) \rightarrow L^1(\mathcal{B}, \mu)$ is a bounded projection onto the subspace $L^1(\mathcal{C}, \mu) \subset L^1(\mathcal{B}, \mu)$ of all \mathcal{C} -measurable and integrable functions.

Let be $p \in [1, \infty]$ and $S = \{\mathbb{1}_U : U \in \mathcal{B}, \mu(U) < \infty\}$. The set S is dense in $L^p(\mathcal{B}, \mu)$ because (X, \mathcal{B}, μ) is σ -finite and $S \subset L^1(\mathcal{B}, \mu) \cap L^p(\mathcal{B}, \mu)$. For all $\mathbb{1}_U \in S$ we have from [] that $\|P_{\mathcal{C}}\mathbb{1}_U\|_p \leq \|\mathbb{1}_U\|_p$ where $\|\cdot\|_p$ denotes the L^p -norm. Hence, $P_{\mathcal{C}}$ extends to a bounded operator

$$P_{\mathcal{C}}^p : L^p(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu). \quad (2.3)$$

Proposition 2.1. Suppose $1 \leq p \leq \infty$. The operator $P_{\mathcal{C}}^p$ is a projection onto the subspace $L^p(\mathcal{C}, \mu)$ of \mathcal{C} -measurable L^p -functions. Furthermore, for all $U \in \mathcal{B}$ with $\mu(U) < \infty$ we have

$$P_{\mathcal{C}}^p(\mathbb{1}_U) = P_{\mathcal{C}}(\mathbb{1}_U)$$

where $P_{\mathcal{C}}(\mathbb{1}_U)$ denotes the Radon-Nikodym derivative of the characteristic function $\mathbb{1}_U$.

Proof. The equality $P_{\mathcal{C}}^p(\mathbb{1}_U) = P_{\mathcal{C}}(\mathbb{1}_U)$, for all $U \in \mathcal{B}$ with $\mu(U) < \infty$, follows from the definition of the operator $P_{\mathcal{C}}^p$. Let be $f \in L^p(\mathcal{C}, \mu)$. Since (X, \mathcal{C}, μ) is a σ □

Theorem 2.2. *Let $T : \mathcal{D}(T) \rightarrow L^p(\mathcal{B}, \mu)$ an operator and $\mathcal{C} \subset \mathcal{B}$ a sub- σ -algebra of \mathcal{B} . Then, there is a factor $\hat{T} : \mathcal{D}(\hat{T}) \rightarrow L^p(\hat{\mathcal{B}}, \hat{\mu})$ and a bounded, surjective operator $\Pi : L^p(\mathcal{B}, \mu) \rightarrow L^p(\hat{\mathcal{B}}, \hat{\mu})$, with support $L^p(\mathcal{C}, \mu)$, if and only if*

$$P_{\mathcal{C}}^p T = P_{\mathcal{C}}^p T P_{\mathcal{C}}^p.$$

Proof. If: On the space X we define the equivalence relation

$$x \sim y \quad :\Leftrightarrow \quad \text{for all } C \in \mathcal{C} (x \in C \text{ iff } y \in C).$$

We call the equivalence class $[x]_{\mathcal{C}}$ the \mathcal{C} -atom of x . We define $\hat{X} = X / \sim$ as the space of all equivalence classes, with the canonical projection $\pi : X \rightarrow \hat{X}; x \mapsto [x]_{\mathcal{C}}$. The setting

$$\hat{\mathcal{B}} = \left\{ \hat{B} \subset \hat{X} : \pi^{-1}(\hat{B}) \in \mathcal{C} \right\}$$

yields a σ -algebra on \hat{X} and the push forward of μ under π defines a measure on \hat{X} , i.e.,

$$\hat{\mu}(\hat{B}) = \mu(\pi^{-1}(\hat{B})) \quad \text{for all } \hat{B} \in \hat{\mathcal{B}} \quad (2.4)$$

The definition

$$U : L^p(\hat{\mathcal{B}}, \hat{\mu}) \rightarrow L^p(\mathcal{B}, \mu); \hat{f} \mapsto \hat{f} \circ \pi$$

provides an isometry w.r.t. the L^p -norm in both spaces. The function $U(\hat{f}) = \hat{f} \circ \pi$ is \mathcal{C} -measurable for all $\hat{f} \in L^p(\hat{\mathcal{B}}, \hat{\mu})$ due to the definition of \hat{B} . Conversely, let $g \in L^p(\mathcal{B}, \mu)$ be a \mathcal{C} -measurable function and $x, y \in X$ such that $g(x) \neq g(y)$. There are open neighborhoods U_x and U_y of $g(x)$ and $g(y)$, respectively, such that $U_x \cap U_y = \emptyset$. $g^{-1}(U_x)$ and $g^{-1}(U_y)$ are \mathcal{C} -measurable subsets of X such that $g^{-1}(U_x) \cap g^{-1}(U_y) = \emptyset$, with $x \in g^{-1}(U_x)$ and $y \in g^{-1}(U_y)$, and $[x]_{\mathcal{C}} \neq [y]_{\mathcal{C}}$ is proven. Hence the function $\hat{g} : \hat{X} \rightarrow \mathbb{C}; [x]_{\mathcal{C}} \mapsto g(x)$ is well defined. Let be $U \subset \mathbb{C}$ open. Then

$$\pi^{-1}(\hat{g}^{-1}(U)) = (\hat{g} \circ \pi)^{-1}(U) = g^{-1}(U) \in \mathcal{C}$$

and \hat{B} -measurability of \hat{g} is proven. L^p -integrability of \hat{g} with respect to $\hat{\mu}$ follows from the one of g with respect to μ . Thus, we verified

$$U(L^p(\hat{\mathcal{B}}, \hat{\mu})) = L^p(\mathcal{C}, \mu).$$

For all $f \in L^p(\mathcal{C}, \mu)$ let \hat{f} denote the unique element in $L^p(\hat{\mathcal{B}}, \hat{\mu})$ such that $f = \hat{f} \circ \pi$. The definition

$$\Pi(f) = \begin{cases} \hat{f} & \text{if } f \in P_{\mathcal{C}}^p(L^p(\mathcal{B}, \mu)) \\ 0 & \text{if } f \in (1 - P_{\mathcal{C}}^p)(L^p(\mathcal{B}, \mu)) \end{cases} \quad (2.5)$$

provides a norm-decreasing, surjective, linear operator on $L^p(\mathcal{B}, \mu)$ with support $L^p(\mathcal{C}, \mu)$.

Let be $x \in \mathcal{D}(\hat{T}) = \Pi(\mathcal{D}(T))$. Then there is a $y \in \mathcal{D}(T)$ such that $x = \Pi y$. We define

$$\hat{T} : \mathcal{D}(\hat{T}) \rightarrow L^p(\hat{\mathcal{B}}, \hat{\mu}); x \mapsto \Pi T y.$$

The definition does not depend on the choice of y . Let also $x = \Pi z$, then $0 = \Pi(z - y)$ and thus $P_{\mathcal{C}}^p(z - y) = 0$ because the support of Π is equal the range of $P_{\mathcal{C}}^p$. This yields

$$\Pi T(z - y) = \Pi P_{\mathcal{C}}^p T(z - y) = \Pi P_{\mathcal{C}}^p T P_{\mathcal{C}}^p(z - y) = 0.$$

Furthermore, for all $z \in \mathcal{D}(T)$ we have

$$\Pi T z = \Pi P_{\mathcal{C}}^p T z = \Pi P_{\mathcal{C}}^p T P_{\mathcal{C}}^p z = \Pi T P_{\mathcal{C}}^p z = \hat{T} \Pi P_{\mathcal{C}}^p z = \hat{T} \Pi z.$$

Only If: From the commutativity relation follows

$$\Pi P_{\mathcal{C}}^p T = \Pi T = \hat{T} \Pi = \hat{T} \Pi P_{\mathcal{C}}^p = \Pi T P_{\mathcal{C}}^p$$

and therefore $\Pi(TP_{\mathcal{C}}^p - P_{\mathcal{C}}^p T) = 0$. Since the domain of Π concurs with $L^p(\mathcal{C}, \mu)$ this yields

$$0 = P_{\mathcal{C}}^p(TP_{\mathcal{C}}^p - P_{\mathcal{C}}^p T) = P_{\mathcal{C}}^p T P_{\mathcal{C}}^p - P_{\mathcal{C}}^p T.$$

□

The condition $P_{\mathcal{C}}^p T = P_{\mathcal{C}}^p T P_{\mathcal{C}}^p$ does not tell us that the subspace of \mathcal{C} -measurable functions in $L^p(\mathcal{B}, \mu)$ is invariant under the operator T . Invariance would require the condition $TP_{\mathcal{C}}^p = P_{\mathcal{C}}^p T P_{\mathcal{C}}^p$ which cannot be followed from the first equality in general. But the relation $P_{\mathcal{C}}^p T = P_{\mathcal{C}}^p T P_{\mathcal{C}}^p$ in the space $L^p(\mathcal{B}, \mu)$ forces that $TP_{\mathcal{C}}^p = P_{\mathcal{C}}^p T P_{\mathcal{C}}^p$ holds true in the dual space.

The dual space $L^p(\mathcal{B}, \mu)'$ consists of all bounded linear functionals on $L^p(\mathcal{B}, \mu)$. Let be $p \in [1, \infty)$ and $q \in (1, \infty]$ such that $1/p + 1/q = 1$ where we set $q = \infty$ if $p = 1$. Additionally, we suppose $f \in L^p(\mathcal{B}, \mu)$ and $g \in L^q(\mathcal{B}, \mu)$. From Hölder's inequality follows $fg \in L^1(\mathcal{B}, \mu)$ and

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$

This inequality proves that the linear functional

$$F_g : L^p(\mathcal{B}, \mu) \rightarrow \mathbb{C}; f \mapsto \int f(x)g(x) d\mu(x) \quad (2.6)$$

is bounded with norm $\|F_g\| \leq \|g\|_q$ and an element of the dual space $L^p(\mathcal{B}, \mu)'$, i.e., the space of all bounded linear functionals on $L^p(\mathcal{B}, \mu)$. Since μ is a σ -finite measure, the Radon-Nikodym theorem yields that any bounded functional has this form, and the operator

$$F : L^q(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu)'; g \mapsto F_g$$

is an isometric isomorphism. For $1 < p < \infty$ this isometric isomorphism even proves reflexivity $L^p(\mathcal{B}, \mu)'' \cong L^p(\mathcal{B}, \mu)$.

Let be $T : L^p(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu)$ a bounded operator. Then, for all $\chi \in L^p(\mathcal{B}, \mu)'$ the composition $\chi \circ T$ is again a bounded functional. We call

$$T' : L^p(\mathcal{B}, \mu)' \rightarrow L^p(\mathcal{B}, \mu)'; \chi \mapsto \chi \circ T$$

the *dual* of T which is a bounded operator as well with norm $\|T'\| = \|T\|$

Definition 2.4. Let be $p \in [1, \infty)$ and $q \in (1, \infty]$ such that $1/p + 1/q = 1$ where we set $q = \infty$ if $p = 1$. Suppose $T : L^p(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu)$ is an operator. The composition

$$T^* = F^{-1} \circ T' \circ F$$

yields a bounded operator on $L^q(\mathcal{B}, \mu)$ which we call the *adjoint* of T .

Corollary 2.3. Let T, S and ST bounded operators. Then, the following equalities hold:

1. If $1 < p < \infty$ we have $(T^*)^* = T$
2. $(TS)^* = S^*T^*$

Proof. Follows immediately from the definition of the adjoint operator and reflexivity of L^p -spaces for $p \in (1, \infty)$. □

Theorem 2.4. Let be $p \in [1, \infty)$ and $q \in (1, \infty]$ such that $1/p + 1/q = 1$ where we set $q = \infty$ if $p = 1$. Suppose $T : L^p(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu)$ is a bounded operator, and $\mathcal{C} \subset \mathcal{B}$ a sub- σ -algebra of \mathcal{B} . Then

$$P_{\mathcal{C}}^p T = P_{\mathcal{C}}^p T P_{\mathcal{C}}^p \quad \text{iff} \quad T^* P_{\mathcal{C}}^q = P_{\mathcal{C}}^q T^* P_{\mathcal{C}}^q,$$

i.e., the subspace of \mathcal{C} -measurable functions in L^q is invariant under T^* .

Proof. Due to corollary 2.3 we need to prove $P_{\mathcal{C}}^{p*} = P_{\mathcal{C}}^q$ only. Since both projections are linear, we need to prove identity only on the positive cone of the Banach space $L^p(\mathcal{B}, \mu)$. Hence, throughout the proof we can assume that all functions are non-negative. We start with the observation that for all $f \in L^q(\mathcal{B}, \mu)$, such that $P_{\mathcal{C}}^{p*} f = f$, the following equivalences hold true:

$$\begin{aligned} & P_{\mathcal{C}}^{p*}(f) = f \\ \Leftrightarrow & P_{\mathcal{C}}^{p'} \circ F(f) = F(f) \\ \Leftrightarrow & \left(P_{\mathcal{C}}^{p'} \circ F(f) \right)(g) = F(f)(g) \quad \text{for all } g \in L^p(\mathcal{B}, \mu) \\ \Leftrightarrow & F(f)(P_{\mathcal{C}}^p g) = F(f)(g) \quad \text{for all } g \in L^p(\mathcal{B}, \mu) \\ \Leftrightarrow & \int f P_{\mathcal{C}}^p g \, d\mu = \int f g \, d\mu \quad \text{for all } g \in L^p(\mathcal{B}, \mu) \\ \Leftrightarrow & \int f (1 - P_{\mathcal{C}}^p) g \, d\mu = 0 \quad \text{for all } g \in L^p(\mathcal{B}, \mu). \end{aligned} \tag{2.7}$$

We first prove the inequality $P_{\mathcal{C}}^{p*} \leq P_{\mathcal{C}}^q$, that is, the image of $P_{\mathcal{C}}^{p*}$ is contained in the one of $P_{\mathcal{C}}^q$. We start with $1 < p$: Assume that $f \geq 0$. Since $f \in L^q(\mathcal{B}, \mu)$ we have $g = f^{q/p} \in L^p(\mathcal{B}, \mu)$. Since projections maps non-negative to non-negative functions we have

$$0 \leq (1 - P_{\mathcal{C}}^p)g \leq g.$$

Let us assume that $(1 - P_{\mathcal{C}}^p)g > 0$. There is a measurable set $U \in \mathcal{B}$ such that $\mu(U) > 0$ and $(1 - P_{\mathcal{C}}^p)g(x) > 0$ for almost all $x \in U$. Then also $g|_U > 0$ a.s. and for this reason also $f|_U > 0$ a.s. This implies

$$\int f(1 - P_{\mathcal{C}}^p)g \, d\mu \geq \int_U f(1 - P_{\mathcal{C}}^p)g \, d\mu > 0$$

– a contradiction. Hence, we obtain $(1 - P_{\mathcal{C}}^p)g = 0$, i.e., $P_{\mathcal{C}}^p g = g$ which proves \mathcal{C} -measurability of g , which is equivalent with the one of f , and $P_{\mathcal{C}}^q(f) = f$ is proven.

1 = p: We have $f \in L^\infty(\mathcal{B}, \mu)$ with $f \geq 0$. Since f is \mathcal{C} -measurable if and only if $f + 1 \in L^\infty(\mathcal{B}, \mu)$ is \mathcal{C} -measurable, we can even assume $f > 0$ a.s. Choose a set $W \in \mathcal{B}$ such that $\mu(W) < \infty$, and define $g = f \mathbb{1}_W$. Since f is bounded, we obtain $g \in L^1(\mathcal{B}, \mu)$. An analogous argument as in the case $1 < p < \infty$ leads to $(1 - P_{\mathcal{C}}^1)g > 0$ – a contradiction as well. Hence, for all $W \in \mathcal{B}$ with $\mu(W) < \infty$ we have $P_{\mathcal{C}}^1(f \mathbb{1}_W) = f \mathbb{1}_W$, i.e., $f \mathbb{1}_W$ is \mathcal{C} -measurable.

Since the measure space (X, \mathcal{B}, μ) is assumed to be σ -finite, there is a countable sequence $(U_n)_{n \in \mathbb{N}}$ of sets in \mathcal{B} such that $\mu(U_n) < \infty$ and $U_{n-1} \subset U_n$ for all $n \in \mathbb{N}$ with $\bigcup U_n = X$. Suppose $V \subset \mathbb{R}^+$ is a Borel set. Since $f > 0$ a.s. and $f \mathbb{1}_{U_n}$ is \mathcal{C} -measurable, we have $(f \mathbb{1}_{U_n})^{-1}(V) = f^{-1}(V) \cap U_n \in \mathcal{C}$. From this follows \mathcal{C} -measurability of f because

$$f^{-1}(V) = \bigcup_{n \in \mathbb{N}} f^{-1}(V) \cap U_n \in \mathcal{C} \quad \text{for all Borel sets } V \subset \mathbb{R}^+.$$

$P_{\mathcal{C}}^q \leq P_{\mathcal{C}}^{p*}$: Conversely, we assume that $f \in L^q(\mathcal{B}, \mu)$ is \mathcal{C} -measurable, i.e., $P_{\mathcal{C}}^q(f) = f$. We want to prove that $P_{\mathcal{C}}^{p*}(f) = f$. The equivalences Eq. (2.7) imply that $P_{\mathcal{C}}^{p*}(f) = f$ iff $F_f \circ P_{\mathcal{C}}^p = F_f$ where F_f denotes the linear functional on $L^p(\mathcal{B}, \mu)$ defined by Eq. (2.6). $F_f \circ P_{\mathcal{C}}^p$ and F_f are bounded linear maps. Therefore, $F_f \circ P_{\mathcal{C}}^p = F_f$ needs to be checked only for a subset of $L^p(\mathcal{B}, \mu)$ whose linear span is dense in $L^p(\mathcal{B}, \mu)$. Since the measure space (X, \mathcal{B}, μ) is assumed to be σ -finite, the linear span of the set $\{\mathbb{1}_U : U \in \mathcal{B}, \mu(U) < \infty\}$ is dense in $L^p(\mathcal{B}, \mu)$. Let be $U \in \mathcal{B}$ such that $\mu(U) < \infty$. The characteristic function 1_U is an L^1 -function, hence its

projection $P_{\mathcal{C}}^p \mathbb{1}_U$ onto the space of \mathcal{C} -measurable functions concurs with its Radon-Nikodym derivative and we obtain for every \mathcal{C} measurable L^1 -function f the identity

$$\int f P_{\mathcal{C}}^p \mathbb{1}_U d\mu = \int f \mathbb{1}_U d\mu. \quad (2.8)$$

But the measure space is σ -finite and the \mathcal{C} -measurable L^1 -functions are dense in $L^q(\mathcal{C}, \mu)$ which provides even

$$\int f P_{\mathcal{C}}^p \mathbb{1}_U d\mu = \int f \mathbb{1}_U d\mu. \quad (2.9)$$

for all $f \in L^q(\mathcal{C}, \mu)$. □

3 Finite state spaces

Markov chains provide a prominent example of a time discrete linear dynamics on finite dimensional vector spaces. Let (\mathcal{B}, μ) be a probability space, and consider a time discrete stochastic process $(Y_m)_{m \in \mathbb{N}}$ such that $Y_m : X \rightarrow \{y_1, \dots, y_n\}$. This induces a series of distribution vectors $p_m \in \mathbb{R}^n$ on $\{y_1, \dots, y_n\}$ via $p_m = (p^1 = \mu(Y_m = y_1), \dots, p^n = \mu(Y_m = y_n))$. We call the stochastic process Markovian if

$$\mu(Y_{m+1} = y_{i_{m+1}} | Y_m = y_{i_m}, \dots, Y_0 = y_{i_0}) = \mu(Y_{m+1} = y_{i_{m+1}} | Y_m = y_{i_m}) \quad \text{for all } m \in \mathbb{N}$$

holds. Furthermore, if we assume stationarity, i.e.,

$$\mu(Y_{m+1} = y' | Y_m = y) = \mu(Y_1 = y' | Y_0 = y) \quad \text{for all } m \in \mathbb{N}.$$

Then, one can define a transition kernel which is given by the $n \times n$ matrix

$$T = (T_{ji}) \quad \text{with} \quad T_{ji} = p(Y_1 = y_i | Y_0 = y_j) \quad \text{for all } i, j = 1, \dots, n$$

where $p(y_i | y_j)$ denotes the transition probability between state y_i and y_j after one time step such that

$$p_{m+1} = p_m T \quad \text{for all } m \in \mathbb{N}, \quad (3.1)$$

with left-multiplication of the row vector p_m with the matrix T .

For these dynamical systems it is natural to ask whether one can aggregate the process in order to obtain a further Markovian process. Often, one considers deterministic aggregations where states of the same equivalence class of the state space are lumped together, see [13]. It rises the question how to find valid lumpings where the corresponding aggregated process is again Markovian. Many papers address this problem like, for example, [5], [2], [4], etc. The work of M. N. Jacobi and O. Görnerup in [14] provides the most general solution to this problem. They use level sets of vectors which are invariant under the transition matrix T , for instance eigenvectors. An example demonstrates the mechanism.

Example 3.1. Consider a stationary Markov process $Y_m : (\mathcal{B}, \mu) \rightarrow \{y_1, \dots, y_4\}$, with $m \in \mathbb{N}$ and transition kernel

$$T = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Diagonalization of the matrix T yields a diagonal matrix $D = \text{diag}(0, 1, -i/2, i/2)$ with **right**-eigenvectors

$$\begin{pmatrix} -1 & 1 & -1 - i & -1 + i \\ 1 & 1 & 2i & -2i \\ 0 & 1 & -1 - i & -1 + i \\ 0 & 1 & 1 & 1 \end{pmatrix}. \quad (3.2)$$

If one has a look at the the matrix Eq. (3.2), one recognizes that the first and the third entry of all eigenvectors, excluding the first one, agree. If one lumps together the first and the third state, there are as many states left as eigenvectors exist whose first and third entry agree. In [14] the authors prove that under these circumstances the aggregated process is Markovian again. Indeed, the lumping described by the aggregation matrix

$$\Pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

leads to a Markovian process whose transition kernel is given by

$$\hat{T} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \end{pmatrix},$$

and one checks easily that $\Pi\hat{T} = T\Pi$, that is, for every distribution vector $p \in \mathbb{R}^n$ we obtain

$$p\Pi\hat{T} = pT\Pi.$$

The example illustrates one crucial point. If one wants to find possible aggregations of a stationary Markov process which can be described by a left multiplication of a distribution vector $p \in \mathbb{R}^n$ with a matrix T , see Eq. (3.1), one needs to investigate the level sets of column vectors which are invariant under right-multiplication with the matrix T . Since left-multiplication of row vectors with T is the same as right-multiplication of column vectors with T^* , the conjugate transposed of T , the close link of the eigenvector method worked out in detail in [14] with our theorems 2.2 and 2.4 is striking. In the sequel we derive the results in [14] from our general framework developed in the previous section.

Let be $n \in \mathbb{N}$ and consider the linear space \mathbb{C}^n , with a linear map $T : \mathbb{C}^n \rightarrow \mathbb{C}^n$ given as $n \times n$ matrix and right multiplication, i. e. $v \mapsto Tv$ for all $v \in \mathbb{C}^n$. We define $X = \{1, \dots, n\}$ and \mathcal{B} as the σ -algebra of all subsets of X , with measure $\mu(i) = 1$ for all $i = 1, \dots, n$. Let be $p \in [1, \infty)$. Then,

$$\phi^p : \mathbb{C}^n \rightarrow L^p(\mathcal{B}, \mu); \quad v = (v^1, \dots, v^n)^T \mapsto f := \{i \mapsto v^i\} \quad (3.3)$$

is an isometric isomorphism with respect to the norm

$$\|v\|_p = \left(\sum_{i=1}^n |v^i|^p \right)^{1/p} \quad \text{for all } v \in \mathbb{C}^n.$$

Then $S = \phi^p \circ T \circ (\phi^p)^{-1}$ yields a bounded operator on $L^p(\mathcal{B}, \mu)$ whose adjoint is given by $\phi^q \circ T^* \circ (\phi^q)^{-1}$ where T^* denotes the conjugate transposed of the matrix T .

Suppose $v_1, \dots, v_k \in \mathbb{C}^n$ are linearly independent such that

$$T^*v_i \in V = \langle v_1, \dots, v_k \rangle \quad \text{for all } i = 1, \dots, k$$

where $\langle \cdot \rangle$ denotes the linear span of the vectors v_1, \dots, v_k . We call such vectors *invariant* under T^* . Let denote with \mathfrak{S}_n the permutation group on n elements. We define a group action of \mathfrak{S}_n on \mathbb{C}^n

$$\rho : \mathfrak{S}_n \rightarrow GL_n(\mathbb{C}); \quad \rho(\tau)(e_i) = e_{\tau(i)}$$

for all $i = 1, \dots, n$ where $GL_n(\mathbb{C})$ denotes the group of invertible $n \times n$ matrices and e_1, \dots, e_n the standard base vectors of \mathbb{C}^n . Let denote with G_i the stabilizer group of v_i , i.e.,

$$G_i = \{ \tau \in \mathfrak{S}_n : \rho(\tau)(v_i) = v_i \},$$

and the group $G = \bigcap_{i=1}^n G_i$. Then, $\tau \in G$ iff $\rho(\tau)$ acts as the identity on V . On the set of indices $\{1, \dots, n\}$ we define the equivalence relation

$$r \sim t \Leftrightarrow \exists \tau \in G \quad \tau(r) = t.$$

We have $t \in [r]$ if and only if $v_i^r = v_i^t$ for all $i = 1, \dots, k$. We call the set of equivalence classes $X/\sim = \{[r] : r = 1, \dots, n\}$ the *level sets* of the family v_1, \dots, v_k . X/\sim is a partition of the set X and we denote with \mathcal{C} the sub- σ -algebra of \mathcal{B} generated by X/\sim . A function $f \in L^q(\mathcal{B}, \mu)$ is \mathcal{C} -measurable if and only if it is constant on the level sets.

Lemma 3.1. *Let be $Q : \mathbb{C}^n \rightarrow \mathbb{C}^n$ the projection onto the subspace $V = \langle v_1, \dots, v_k \rangle$. Then, the inequality $Q \leq (\phi^q)^{-1} \circ P_{\mathcal{C}}^q \circ \phi^q$ holds.*

Proof. Let $v \in \mathbb{C}^n$ with $Qv = v$. There are $\lambda_i \in \mathbb{C}$ such that

$$v = \sum_{i=1}^k \lambda_i v_i$$

Let $r \in \{1, \dots, n\}$ and $t \in [r]$ then $v_i^r = v_i^t$ for all $i = 1, \dots, k$ which provides also $v^r = v^t$. Therefore the function $\phi^q(v)$ is constant on the level sets. This proves \mathcal{C} -measurability of $\phi^q(v)$. This yields

$$(\phi^q)^{-1} \circ P_{\mathcal{C}}^q \circ \phi^q(v) = (\phi^q)^{-1} (\phi^q(v)) = v$$

□

Corollary 3.2. *The number k of linearly independent invariant vectors of T^* is in general smaller than the number of level sets they generate.*

Proof. $k = \text{rank } Q$, the rank of the projection Q , and the number of level sets is equal the rank of $P_{\mathcal{C}}^p$. Application of lemma 3.1 proves the corollary. □

Let be the number k of linearly independent vectors which are invariant under T^* equal the number of level sets they generate. Then $Q = (\phi^q)^{-1} \circ P_{\mathcal{C}}^q \circ \phi^q$. From $T^*Q = QT^*Q$ follows also $S^*P_{\mathcal{C}}^q = P_{\mathcal{C}}^q S^*P_{\mathcal{C}}^q$, i.e., not only the k linearly independent vectors v_1, \dots, v_k are invariant under T^* but also the space of all \mathcal{C} -measurable functions is invariant under S^* .

From theorem 2.4 and 2.2 follows the existence of a factor \hat{S} . Having a look at the proof of theorem 2.2, we are even able to work out the precise form of its corresponding aggregation $\Pi : L^p(X, \mu) \rightarrow L^p(\hat{\mathcal{B}}, \hat{\mu})$. The proof provides $\hat{X} = X/\sim$, $\hat{\mu}$ is given by Eq. (2.4), and $\hat{\mathcal{B}}$ is the set of all subsets of X/\sim . More subtle is the description of the linear map Π . Let us define

$$\Lambda(f)([r]) = \frac{1}{|[r]|} \sum_{t \in [r]} f(t) \quad \text{for all } f \in L^p(\mathcal{B}, \mu). \quad (3.4)$$

In the proof of 2.2 we saw that $\hat{f} \in L^p(\hat{\mathcal{B}}, \hat{\mu})$ if and only if there is a unique \mathcal{C} -measurable function $f \in L^p(\mathcal{B}, \mu)$ such that $\hat{f} = f \circ \pi$ where $\pi : X \rightarrow X/\sim; r \mapsto [r]$ denotes the canonical projection. From Eq. (3.4) follows $\Lambda(f) = \hat{f}$. Hence Λ maps the space $L^p(\mathcal{C}, \mu)$ injectively onto $L^p(\hat{\mathcal{B}}, \hat{\mu})$. From the dimension formula for linear maps on finite linear spaces follows $\Lambda(1 - P_{\mathcal{C}}^p) = 0$ and therefore $\Lambda = \Pi$. The factor $\hat{S} : L^p(\hat{\mathcal{B}}, \hat{\mu}) \rightarrow L^p(\hat{\mathcal{B}}, \hat{\mu})$ is given by $\hat{f} \mapsto \Pi S(f)$ where $f \in L^p(\mathcal{C}, \mu)$ denotes the unique \mathcal{C} -measurable function such that $\hat{f} = f \circ \pi$. Let $\varphi : \{1, \dots, k\} \rightarrow X/\sim$ be a bijection which induces a total order on X/\sim . Define for $1 \leq p < \infty$

$$\theta^p : \mathbb{C}^k \rightarrow L^p(\hat{\mathcal{B}}, \hat{\mu}); \quad v = (v^1, \dots, v^k)^T \mapsto f := \{[r] \mapsto v^{\varphi^{-1}([r])}\}.$$

Similar as before, this induces an isometric isomorphism between $L^p(\hat{\mathcal{B}}, \hat{\mu})$ and \mathbb{C}^k with the slight difference that the norm on \mathbb{C}^k has the form

$$\|v\|_p = \left(\sum_{i=1}^k |v^i|^p |\varphi(i)| \right)^{1/p} \quad \text{for all } v = (v^1, \dots, v^k) \in \mathbb{C}^k$$

where $|\varphi(i)|$ means the cardinality of the level set $\varphi(i)$. The composition $(\theta^p)^{-1} \circ \Pi \circ \phi^p : \mathbb{C}^n \rightarrow \mathbb{C}^k$ leads to a $k \times n$ aggregation matrix Π_T with entries

$$\pi_{ij} = \begin{cases} \frac{1}{|\varphi(i)|} & \text{if } \varphi(i) = [r], \text{ and } j \in [r] \\ 0 & \text{else} \end{cases} . \quad (3.5)$$

From Eq. (3.5) one reads off that the rows of the Π_T are invariant under the action of the group G , i.e., we have $\pi_{i\rho(\tau)(j)} = \pi_{ij}$ for all $i \in \{1, \dots, k\}$, $j \in \{1, \dots, n\}$ and $\tau \in G$. If we define $\hat{T} = (\theta^p)^{-1} \circ \hat{S} \circ \theta^p$, we attain the commutative diagram

$$\begin{array}{ccc} \mathbb{C}^n & \xrightarrow{T} & \mathbb{C}^n \\ \Pi_T \downarrow & & \downarrow \Pi_T \\ \mathbb{C}^k & \xrightarrow{\hat{T}} & \mathbb{C}^k \end{array} \quad (3.6)$$

Corollary 3.3. *Let T be an $n \times n$ transition matrix, describing a linear map $T : \mathbb{C}^n \rightarrow \mathbb{C}^n$. Then, for a $k \times n$ aggregation matrix Π_T of the form Eq. (3.5), with $k \leq n$, exists a $k \times k$ matrix \hat{T} , such that the diagram 3.6 commutes if and only if there is a set of k -linearly independent vectors v_1, \dots, v_k , invariant under T^* , whose level sets are invariant under the same permutations as the rows of Π_T .*

Proof. We need to prove the ‘‘only if’’ direction. Suppose there is an aggregation matrix of the form Eq. (3.5) and a $k \times k$ matrix \hat{T} such that $\Pi_T T = \hat{T} \Pi_T$. Then also $T^* \Pi_T^t = \Pi_T^t \hat{T}^*$ where A^t denotes the transpose of the matrix A . But the last equation tells us that the k columns of Π_T^t , which are the rows of Π , are linearly independent and invariant under T^* . \square

Remark 3.1. Corollary 3.3 was proven by M. N. Jacobi and O. Görnerup in [14]. We reproduced their result within our infinite setting which is a proper generalization of their approach working only for time discrete dynamics on finite spaces.

4 Point Transformations

Let be (X, \mathcal{B}, μ) a σ -finite measure space and $\varphi : X \rightarrow X$ a measurable and measure preserving transformation, i.e., for all $B \in \mathcal{B}$ we have $\mu(\varphi^{-1}(B)) = \mu(B)$. φ induces an operator $U_\varphi : L^p(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu)$; $f \mapsto f \circ \varphi$, which is an isometry because φ is measure preserving. This yields

$$\|f \circ \varphi\|_p^p = \int |f \circ \varphi|^p d\mu = \int |f|^p d\mu = \|f\|_p^p.$$

in the case $1 \leq p < \infty$. If $p = \infty$, then $\|f \circ \varphi\|_\infty \leq \|f\|_\infty$ which yields $\|U_\varphi\|_\infty \leq 1$ and U_φ is norm-decreasing.

Proposition 4.1. *If $\varphi^{-1}(\mathcal{C}) \subset \mathcal{C}$, then the adjoint $U_\varphi^* : L^q(\mathcal{B}, \mu) \rightarrow L^q(\mathcal{B}, \mu)$, with $1 < q \leq \infty$, has a factor \widehat{U}_φ^* with respect to the sub- σ -algebra \mathcal{C} and $\widehat{U}_\varphi^* = U_\varphi^*$, with $\hat{\varphi} : \hat{X} \rightarrow \hat{X}$ given by*

$$[x]_{\mathcal{C}} \mapsto [\varphi(x)]_{\mathcal{C}} \quad (4.1)$$

for all \mathcal{C} -atoms $[x]_{\mathcal{C}}$.

Conversely, if there is a factor of the adjoint U_φ^* , and \mathcal{C} is σ -finite then $\varphi^{-1}(\bar{\mathcal{C}}) \subset \bar{\mathcal{C}}$ where $\bar{\mathcal{C}}$ denotes the completion of \mathcal{C} .

Proof. If: Let $y \in [x]_{\mathcal{C}}$. Since $\varphi^{-1}(\mathcal{C}) \subset \mathcal{C}$, for all $C \in \mathcal{C}$ we obtain

$$\varphi(x) \in C \Leftrightarrow x \in \varphi^{-1}(C) \Leftrightarrow y \in \varphi^{-1}(C) \Leftrightarrow \varphi(y) \in C .$$

Hence the map $\hat{\varphi}$, given by Eq. (4.1), is well defined and we obtain the commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{\varphi} & X \\ \pi \downarrow & & \downarrow \pi \\ \hat{X} & \xrightarrow{\hat{\varphi}} & \hat{X} \end{array} \quad (4.2)$$

with the canonical projection $\pi : X \rightarrow \hat{X}; x \mapsto [x]_{\mathcal{C}}$. This induces a commutative diagram

$$\begin{array}{ccc} L^q(\mathcal{B}, \mu) & \xrightarrow{U_{\varphi}^*} & L^q(\mathcal{B}, \mu) \\ \Pi \downarrow & & \downarrow \Pi \\ L^q(\hat{\mathcal{B}}, \hat{\mu}) & \xrightarrow{U_{\hat{\varphi}}^*} & L^q(\hat{\mathcal{B}}, \hat{\mu}) \end{array} . \quad (4.3)$$

for $1 < q \leq \infty$. From theorem 2.2 surjectivity of Π follows and this forces the factor to be unique. Thus we obtain $U_{\hat{\varphi}}^* = \widehat{U_{\varphi}^*}$.

Only if: $P_{\mathcal{C}}^p$ is a projection onto the space of \mathcal{C} -measurable functions of $L^p(\mathcal{B}, \mu)$. Theorems 2.4 and 2.2 provide

$$\begin{aligned} U_{\varphi}(f) &\in L^p(\mathcal{C}, \mu) \quad \text{for all } f \in L^p(\mathcal{C}, \mu) \\ \Leftrightarrow f \circ \varphi &\in L^p(\mathcal{C}, \mu) \quad \text{for all } f \in L^p(\mathcal{C}, \mu) \end{aligned} \quad (4.4)$$

for $p \in [1, \infty)$. Since \mathcal{C} is assumed to be σ -finite, there is an increasing sequence $\{U_n\}_{n \in \mathbb{N}}$ of \mathcal{C} -measurable sets with finite measure whose union is a.s. X . Consider a set $C \in \mathcal{C}$. Define $C_n = C \cap U_n \in \mathcal{C}$ and

$$f = \sum_{n \in \mathbb{N}} 2^{-n} \mathbb{1}_{C_n} \in L^p(\mathcal{C}, \mu).$$

From the monotone convergence theorem follows $f^{-1}(\mathbb{R}^+) = C$ a.s. Hence, $\varphi^{-1}(C) = (f \circ \varphi)^{-1}(\mathbb{R}^+)$ a.s., and from Eq. (4.4) we obtain $\varphi^{-1}(C) \in \bar{\mathcal{C}}$. This yields $\varphi^{-1}(C) \subset \bar{\mathcal{C}}$. Since φ is measure preserving, we have $\mu(\varphi^{-1}(N)) = \mu(N) = 0$ for all sets N of measure zero and $\varphi^{-1}(\bar{\mathcal{C}}) \subset \bar{\mathcal{C}}$ is proven. \square

Example 4.1. In the sequel let be X a space with σ -algebra \mathcal{B} and a measurable map $\varphi : X \rightarrow X$.

(1) Consider an additional space Y with σ -algebra \mathcal{D} and measurable maps $S : X \rightarrow X$, $\varphi : Y \rightarrow Y$, and their product $S \times \varphi : X \times Y \rightarrow X \times Y$. Then \mathcal{B} and \mathcal{D} can be considered as sub- σ -algebras \mathcal{B}' and \mathcal{D}' of $\mathcal{B} \otimes \mathcal{D}$ in terms of $A \mapsto A \times Y$ and $B \mapsto X \times B$. We have

$$(S \times \varphi)^{-1}(\mathcal{B}') \subset \mathcal{B}', \quad (S \times \varphi)^{-1}(\mathcal{D}') \subset \mathcal{D}'.$$

(2) Let be $x, y \in X$ with $x \neq y$ and assume that there are elements $B_1, B_2 \in \mathcal{B}$ with $x \in B_1$ and $y \in B_2$ fulfilling $B_1 \cap B_2 = \emptyset$. Suppose A is a φ -invariant measurable set, that is $\varphi^{-1}(A) = A$. Then the complement of A is also φ -invariant because $\varphi^{-1}(X \setminus A) = \varphi^{-1}(X) \setminus \varphi^{-1}(A) = X \setminus A$. We consider the σ -algebra

$$\mathcal{B}_A := \{A \cap B : B \in \mathcal{B}\} \cup \{X \setminus A\}.$$

This implies

$$[x]_{\mathcal{B}_A} = \begin{cases} \{x\}, & \text{if } x \in A \\ X \setminus A, & \text{if } x \in X \setminus A. \end{cases}$$

That is, the set \hat{X} of \mathcal{B}_A -atoms can be identified with A and an additional point. The dynamics is then given by

$$\hat{\varphi} : \hat{X} \rightarrow \hat{X}, \quad [x]_{\mathcal{B}_A} \mapsto \begin{cases} \{\varphi(x)\}, & \text{if } x \in A \\ X \setminus A, & \text{if } x \in X \setminus A. \end{cases}$$

(3) Consider a linear subspace V of the space $L^0(X, \mathcal{B})$ of measurable \mathbb{C} -valued functions on X . We define the corresponding σ -algebra $\sigma(V)$ as the one which is generated by all sets

$$f^{-1}(U) \quad \text{for all } U \subset \mathbb{C} \text{ Borel measurable, } f \in V.$$

Let U_φ be the operator on $L^0(X, \mathcal{B})$ given by $f \mapsto f \circ \varphi$ and assume $U_\varphi(V) \subset V$. Then

$$\varphi^{-1}(V) \subset \sigma(V)$$

which can be seen as follows. The inverse image of a map commutes with the set operations. Hence the σ -algebra $\varphi^{-1}(\sigma(V))$ is generated by the sets

$$\varphi^{-1}(f^{-1}(U)) = (f \circ \varphi)^{-1}(U) \quad \text{for all } U \subset \mathbb{C} \text{ Borel measurable, } f \in V.$$

Since $f \circ \varphi = U_\varphi(f) = g$ for some $g \in V$, we have $(f \circ \varphi)^{-1}(U) = g^{-1}(U) \in \sigma(V)$, and the inclusion is proven.

(4) Further examples can be constructed by the following procedure. Let be $C \in \mathcal{B}$. Then

$$\mathcal{C} := \bigvee_{n=0}^{\infty} \varphi^{-n}(C),$$

with $\varphi^0 = \text{id}_X$, denotes the σ -algebra generated by the sets $\varphi^{-n}(C)$. It is the smallest sub- σ -algebra of \mathcal{B} that satisfies $C \in \mathcal{C}$ and $\varphi^{-1}(\mathcal{C}) \subset \mathcal{C}$.

Suppose there is a family $\{f_\iota\}_{\iota \in I}$ of eigenfunctions of the operator $U_\varphi : L^p(\mathcal{B}, \mu) \rightarrow L^p(\mathcal{B}, \mu)$, i.e., $f_\iota \in L^p(\mathcal{B}, \mu)$, $f_\iota \neq 0$, and $U_\varphi f_\iota = \lambda_\iota f_\iota$ for an $\lambda_\iota \in \mathbb{C}$. We denote with $\sigma(f_\iota : \iota \in I)$ the σ -algebra generated by the eigenfunctions $\{f_\iota\}_{\iota \in I}$, that is, generated by the sets

$$f_{\lambda_\iota}^{-1}(U) \quad \text{for all } U \subset \mathbb{C} \text{ Borel measurable.} \quad (4.5)$$

Lemma 4.2. *Suppose there is a family $\{f_\iota\}_{\iota \in I}$ of eigenfunctions of the operator U_φ . Then, the σ -algebra $\sigma(f_\iota : \iota \in I)$ is invariant under φ , i.e.,*

$$\varphi^{-1}(\sigma(f_\iota : \iota \in I)) \subset \sigma(f_\iota : \iota \in I).$$

Proof. Since the inverse of a map commutes with the set operations, the σ -algebra $\varphi^{-1}(\sigma(f_\iota : \iota \in I))$ is generated by the sets

$$\varphi^{-1}(f_\iota^{-1}(U)) = (f_\iota \circ \varphi)^{-1}(U) = (\lambda_\iota f_\iota)^{-1}(U) = f_\iota^{-1}(M_{\lambda_\iota}(U))$$

where $U \subset \mathbb{C}$ denotes a Borel measurable subset and $M_{\lambda_\iota} : \mathbb{C} \rightarrow \mathbb{C}; z \mapsto \lambda_\iota z$ the map given by multiplication with λ_ι . If $\lambda_\iota = 0$, we have $\varphi^{-1}(f_\iota^{-1}(U)) \in \{X, \emptyset\}$, depending whether $0 \in U$ or not. If $\lambda_\iota \neq 0$, then not only M_{λ_ι} is measurable but also its inverse $M_{\lambda_\iota}^{-1} = M_{\lambda_\iota^{-1}}$. Hence, U is Borel measurable if and only if this holds for $M_{\lambda_\iota}(U)$. \square

Corollary 4.3. *Suppose there is a family $\{f_\iota\}_{\iota \in I}$ of eigenfunctions of the operator U_φ . Then, the σ -algebra $\sigma(f_\iota : \iota \in I)$ generated by the set Eq. (4.5) defines a factor of of the adjoint operator U_φ^* .*

Proof. Follows immediately form lemma 4.2 and proposition 4.1. \square

Hence, eigenfunctions $\{f_\iota\}_{\iota \in I}$ of the induced operator U_φ are a useful tool for finding sub- σ -algebras which are invariant under the map $\varphi : X \rightarrow X$, see lemma 4.2, and therefore for finding a factor of the adjoint U_φ^* . The eigenfunction method is closely related to example 4.1(3). Define the linear subspace $V = \overline{\langle f_\iota : \iota \in I \rangle}^{\|\cdot\|_p}$, where $\overline{\langle \cdot \rangle}^{\|\cdot\|_p}$ denotes the closure of the linear subspace of $L^p(\mathcal{B}, \mu)$ spanned by the the

eigenfunctions $\{f_\iota\}$ w.r.t. the L^p -norm. Let be $\sigma(V)$ the σ -algebra of example 4.1(3). For every function $f \in V$ there is a sequence of $(f_n)_{n \in \mathbb{N}}$ of functions in $\langle f_\iota : \iota \in I \rangle$ which converges to f w.r.t. the L^p -norm. But every f_n is a finite linear combination of eigenfunctions $\{f_\iota\}_{\iota \in I}$ and therefore $\sigma(f_\iota : \iota \in I)$ -measurable. From the dominated convergence theorem follows also the $\sigma(f_\iota : \iota \in I)$ -measurability of f . This proves $\sigma(V) = \sigma(f_\iota : \iota \in I)$, and the eigenfunction method is therefore a useful tool to find invariant subspaces of the operator U_φ whose corresponding σ -algebra $\sigma(V)$ induces a factor.

From lemma 4.2 we know that every family of eigenfunctions $\{f_\iota\}_{\iota \in I}$ gives us a σ -algebra which is invariant under φ . But in general there are sub- σ -algebras which cannot be obtained in this way.

Example 4.2. Let $k \geq 2$ be a fixed integer, $p = (p_0, p_1, \dots, p_{k-1})$ be a probability vector with non-zero entries (i.e., $p_i > 0$ for each i). $(Y, 2^Y, \mu)$ denote the measure space where $Y = \{0, 1, \dots, k-1\}$ and the point i has measure p_i . Let $X = \prod_{-\infty}^{\infty} X$ be the infinite product, and \mathcal{B} the σ -algebra which is generated by all cylinder sets

$$(a_m, \dots, a_n) = \{x \in X : x_i = a_i \text{ for } i = m, \dots, n\},$$

for $m \leq n$, $m, n \in \mathbb{Z}$, and $a_i \in Y$, where x_i denotes the i 'th component of x . We put

$$\mu(a_0, \dots, a_n) = p_{a_0} p_{a_1} p_{a_2} \cdots p_{a_n}$$

which defines a probability measure on X . Then one can check that $\varphi : X \rightarrow X$ by $\varphi(\{x_n\}) = \{x'_n\}$ where $x'_n = x_{n+1}$ defines a measure preserving map. We call φ the two-sided $(p_0, p_1, \dots, p_{k-1})$ -Markov shift.

Let for our purposes be $k = 3$ and $p_0 = p_1 = p_2 = 1/3$. We define a partition $\mathfrak{A} = \{A_1, A_2\}$ of X setting $A_1 = \{x : x_0 \in \{1, 2\}\}$ and $A_2 = \{x : x_0 = 3\}$. We define

$$\bigvee_{n=-\infty}^{\infty} \varphi^n \mathfrak{A} \tag{4.6}$$

as the sub- σ -algebra of \mathcal{B} which is generated by $\{\{\varphi^n \mathfrak{A}\} : n \in \mathbb{Z}\}$, i.e., it is the intersection of all those sub- σ -algebras of \mathcal{B} that contain all $\varphi^n \mathfrak{A}$. We want to compute the entropy

$$h(\varphi, \mathfrak{A}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\mathfrak{A} \vee \varphi^{-1} \mathfrak{A} \vee \cdots \vee \varphi^{-n+1} \mathfrak{A}).$$

A typical element of $\mathfrak{A} \vee \varphi^{-1} \mathfrak{A} \vee \cdots \vee \varphi^{-n+1} \mathfrak{A}$ is

$$\begin{aligned} & A_{i_0} \cap \varphi^{-1} A_{i_1} \cap \cdots \cap \varphi^{-n+1} A_{i_{n-1}} = \\ & = \{x \in X : x_0 \in A_{i_0}, x_1 \in A_{i_1}, \dots, x_{n-1} \in A_{i_{n-1}}\} \end{aligned}$$

which has measure $1/3^{r_i} \cdot 2/3^{s_i}$ where s_i denotes the number of 1's in the binary sequence $i = (i_0, \dots, i_{n-1})$ and $r_i = n - s_i$. Thus,

$$\begin{aligned} & H(\mathfrak{A} \vee \varphi^{-1} \mathfrak{A} \vee \cdots \vee \varphi^{-n+1} \mathfrak{A}) = \\ & = - \sum_{i \in \{0,1\}^n} 1/3^{r_i} \cdot 2/3^{s_i} \log(1/3^{r_i} \cdot 2/3^{s_i}) \\ & = -n(1/3 \log(1/3) + 2/3 \log(2/3)). \end{aligned}$$

Therefore, $h(\varphi, \mathfrak{A}) = 1/3 \log(3) + 2/3 \log(3/2) < \log(3) = h(\varphi)$, where $h(\varphi)$ denotes the Kolmogorov-Sinai entropy of φ . From the Kolmogorov-Sinai theorem, see theorem 4.17, p.95 in [18], we obtain that Eq. (4.6) defines a proper φ -invariant sub- σ -algebra of \mathcal{B} .

Due to theorem 1.30, p. 51 in [18], the system $\varphi : X \rightarrow X$ is strongly mixing. From this and theorem 1.26, p. 48 in [18], we obtain that the only eigenfunctions of U_φ in $L^2(\mathcal{B}, \mu)$ are the constant functions. Hence, the φ -invariant sub- σ -algebra Eq. (4.6) cannot be generated by any set of eigenfunctions.

The example 4.1(3) is a generalization of one direction of proposition 3.3: If there are k linearly independent vectors which are invariant under φ^* and induce k -level sets, then a lumping Eq. (3.5) can be performed. We can prove an even stronger result by getting rid of the restriction finding as many linearly independent invariant vectors as level sets they induce.

Corollary 4.4. *Let $T : \mathbb{C}^n \rightarrow \mathbb{C}^n$ be a linear map such that $Te_j = e_{\nu(j)}$, for $j = 1, \dots, n$, where $e_i \in \mathbb{C}$ denotes the standard base of \mathbb{C}^n and $\nu : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ a (not necessarily bijective) map.*

Suppose v_1, \dots, v_m are vectors invariant under T^ . Then there is a $k \times n$ aggregation matrix Π_T of the form Eq. (3.5), with $k \leq n$, and a $k \times k$ matrix \hat{T} , such that the diagram 3.6 commutes, and the rows of Π are invariant under the same permutations as the level sets of the vectors v_1, \dots, v_m .*

Proof. Let $S = \varphi^p \circ T^* \circ (\varphi^p)^{-1}$ where $\varphi^p : \mathbb{C}^n \rightarrow L^p(\mathcal{B}, \mu)$ is given by Eq. (3.3), with $X = \{1, \dots, n\}$ and the measure $\mu(j) = 1$ for $j = 1, \dots, n$. We define the function $\delta_j : X \rightarrow \{0, 1\}$, $k \mapsto \delta_{kj}$, with the Kronecker delta δ_{kj} . The set $\{\delta_j : j = 1, \dots, n\}$ is a base of $L^p(\mathcal{B}, \mu)$, and we have

$$S\delta_j = \varphi^p \circ T^* \circ (\varphi^p)^{-1}(\delta_j) = \varphi^p(T^*e_j) = \sum_{m \in \nu^{-1}(j)} \delta_m = \delta_j \circ \nu$$

which proves that S is induced by the transformation $\nu : X \rightarrow X$. Application of proposition 3.3, proposition 4.1 and example 4.1(3) gives us the result. \square

Suppose $\varphi : X \rightarrow X$ is not only measure preserving but has also a measurable inverse φ^{-1} . For all $B \in \mathcal{B}$ we have $\mu(\varphi(B)) = \mu(\varphi^{-1}(\varphi(B))) = \mu(B)$ which proves that φ^{-1} is also measure preserving. For invertible maps φ the adjoint of U_φ is induced by a point transformation again – namely φ^{-1} .

For all $f \in L^p(\mathcal{B}, \mu)$, with $1 \leq p < \infty$, we have $U_\varphi^*(f) = F^{-1} \circ U'_\varphi \circ F_f$, where U'_φ denotes the dual of U_φ , and $F : L^p(\mathcal{B}, \mu) \rightarrow L^q(\mathcal{B}, \mu)'$ is the map Eq. (2.6), with $1 < q \leq \infty$, such that $1/p + 1/q = 1$. For all $g \in L^q(\mathcal{B}, \mu)$ we have

$$U'_\varphi \circ F_f(g) = F_f(U_\varphi(g)) = \int fg \circ \varphi d\mu = \int f \circ \varphi^{-1}g d\mu = F_{f \circ \varphi^{-1}}(g)$$

which proves $U_\varphi^*(f) = f \circ \varphi^{-1} = U_{\varphi^{-1}}(f)$ for all $f \in L^p(\mathcal{B}, \mu)$, hence

$$U_\varphi^* = U_{\varphi^{-1}}, \quad (4.7)$$

an equation which provides an extension of the definition of an adjoint operator to the case $p = 1$. The identification of U_φ^* with $U_{\varphi^{-1}}$ requires that $\hat{\varphi}$ is invertible which does not hold true in general. Clearly, $\hat{\varphi}$ is surjective but it may fail to be injective, a phenomenon which does not occur if the set X is finite, see [16].

Example 4.3. Suppose $\varphi : X \rightarrow X$ is a two-sided Markov shift, defined in example 4.2, with $k = 2$ and the probability vector $p_0 = p_1 = 1/2$. We define a partition $\mathcal{A} = \{A_1, A_2\}$ of X setting $A_1 = \{x : x_0 = 0\}$ and $A_2 = \{x : x_0 = 1\}$. We define

$$\mathcal{C} = \bigvee_{n=0}^{\infty} \varphi^{-n}(\mathcal{A})$$

as the sub- σ -algebra of \mathcal{B} which is generated by all sets $\varphi^{-n}(\mathcal{A})$ with $n \geq 0$ and $\varphi^0 = \text{id}_X$. Let be $x^1, x^2 \in X$ such that $x_n^1 = 0$ for all $n \in \mathbb{Z}$, and $x_n^2 = 0$ for all $n \leq -1$ and $x_n^2 = 1$ for all $n \geq 0$. By definition of \mathcal{C} we have $[x^1]_{\mathcal{C}} \neq [x^2]_{\mathcal{C}}$ because $x^1 \in A_1$ and $x^2 \in A_2$. But $\varphi(x^1)_n = \varphi(x^2)_n$ for all $n \leq 0$ and therefore $[\varphi(x^1)]_{\mathcal{C}} = [\varphi(x^2)]_{\mathcal{C}}$. Hence, $\hat{\varphi}$ is not injective.

If $\hat{\varphi}$ is not injective, for a invertible, measurable and measure preserving map φ with invariant sub- σ -algebra \mathcal{C} , we have $\varphi^{-1}(\mathcal{C}) \neq \mathcal{C}$. Otherwise φ^{-1} keeps the sub- σ -algebra \mathcal{C} also invariant, and the map $[x]_{\mathcal{C}} \mapsto [\varphi^{-1}(x)]_{\mathcal{C}}$ is well defined and injectivity of $\hat{\varphi}$ follows. For finite measure spaces, i.e. $\mu(X) < \infty$, vanishing Kolmogorov-Sinai entropy $h(\varphi) = 0$, is a sufficient condition for $\varphi^{-1}(\overline{\mathcal{C}}) = \overline{\mathcal{C}}$ where $\overline{\mathcal{C}}$ denotes the completion of \mathcal{C} , see corollary 4.14.4 in [18].

Corollary 4.5. Suppose (X, \mathcal{B}, μ) is a finite measure space, φ measurable, measure preserving, invertible with inverse φ^{-1} , and $h(\varphi) = 0$. Let denote with \mathcal{C} a sub- σ -algebra and with $\overline{\mathcal{C}}$ its completion.

If $\varphi^{-1}(\mathcal{C}) \subset \mathcal{C}$, then $\varphi^{-1}(\overline{\mathcal{C}}) = \overline{\mathcal{C}}$, and $U_{\varphi^{-1}} : L^q(\mathcal{B}, \mu) \rightarrow L^q(\mathcal{B}, \mu)$, with $q \in [1, \infty]$, has a factor $\widehat{U_{\varphi^{-1}}}$ with respect to the sub- σ -algebra $\overline{\mathcal{C}}$ and $\widehat{U_{\varphi^{-1}}} = U_{\hat{\varphi}^{-1}}$, with $\hat{\varphi} : \hat{X} \rightarrow \hat{X}$ given by

$$[x]_{\overline{\mathcal{C}}} \mapsto [\varphi(x)]_{\overline{\mathcal{C}}} \quad (4.8)$$

for all $\overline{\mathcal{C}}$ -atoms $[x]_{\overline{\mathcal{C}}}$.

Conversely, if there is a factor of the adjoint U_{φ}^* , and $\overline{\mathcal{C}}$ is σ -finite then $\varphi^{-1}(\overline{\mathcal{C}}) \subset \overline{\mathcal{C}}$.

Proof. Follows immediately from the fact that $\varphi^{-1}(\mathcal{C}) \subset \mathcal{C}$ implies also $\varphi^{-1}(\overline{\mathcal{C}}) \subset \overline{\mathcal{C}}$. \square

5 Markov Kernel

In this section X is a locally compact Hausdorff space, endowed with the Borel σ -algebra \mathcal{B} , and a positive Radon measure μ on X with full support and $\mu(X) = 1$. We consider a Markov kernel $\kappa : X \times \mathcal{B} \rightarrow [0, 1]$, i.e.,

1. $\kappa(\cdot, B)$ \mathcal{B} -measurable for all $B \in \mathcal{B}$
2. $\kappa(x, \cdot)$ probability measure on \mathcal{B} dominated by μ

where dominance means that $\mu(N) = 0$ implies $\kappa(x, N) = 0$. We define the pullback $U_{\kappa} : L^{\infty}(\mathcal{B}, \mu) \rightarrow L^{\infty}(\mathcal{B}, \mu)$ via

$$U_{\kappa}f(x) = \int f(y)\kappa(x, dy).$$

U_{κ} is a linear map with $\|U_{\kappa}\|_{\infty} = 1$ and $U_{\kappa}(1) = 1$ where $1 \in L^{\infty}(\mathcal{B}, \mu)$ denotes the constant function $x \mapsto 1$. The mapping

$$f \mapsto \xi(f) = \int f d\mu \quad (5.1)$$

defines a state on $L^{\infty}(\mathcal{B}, \mu)$, i.e., a bounded linear functional with $\xi(1) = 1$. Since μ has full support, it follows that $\xi(f) > 0$ for all $f > 0$, i.e., ξ is also faithful.

Definition 5.1. We call the Markov kernel κ measure preserving if $\xi \circ U_{\kappa} = \xi$ for the state ξ defined by Eq. (5.1).

From theorem 2, p.59 in [7] we obtain the inequality $U_{\kappa}\overline{f}U_{\kappa}f \leq U_{\kappa}(\overline{f}f)$ for all $f \in L^{\infty}(\mathcal{B}, \mu)$ where \overline{f} denotes the complex conjugate of f . If we assume further that κ is measure preserving, we get

$$\|U_{\kappa}f\|_2^2 = \xi(\overline{U_{\kappa}f}U_{\kappa}f) = \xi(U_{\kappa}\overline{f}U_{\kappa}f) \leq \xi(U_{\kappa}(\overline{f}f)) = \xi(\overline{f}f) = \|f\|_2^2.$$

Thus, U_{κ} is norm-decreasing with respect to the L^2 -norm. Thus, U_{κ} defines a contracting linear map on the dense subspace $L^{\infty}(\mathcal{B}, \mu)$ of $L^2(\mathcal{B}, \mu)$ and can be extended uniquely on the entire space $L^2(\mathcal{B}, \mu)$. We denote the extension of U_{κ} again with U_{κ} and with U_{κ}^* the adjoint operator of U_{κ} on $L^2(\mathcal{B}, \mu)$. Additionally, from Hölder's inequality it follows that every square integrable function $f \in L^2(\mathcal{B}, \mu)$ is also integrable with

$$|\xi(f)| = \left| \int f d\mu \right| \leq \int |f| d\mu \leq \|f\|_2.$$

Thus, ξ can be restricted to a bounded linear functional on $L^2(\mathcal{B}, \mu)$. The definition 2.3 of the adjoint yields

$$\begin{aligned} \xi \circ U_{\kappa}^*(f) &= \xi(1U_{\kappa}^*f) = \int 1U_{\kappa}^*f d\mu = F(U_{\kappa}^*f)(1) \\ &= ((U_{\kappa}' \circ F)(f))(1) = (U_{\kappa}'(Ff))(1) = F_f(U_{\kappa}1) = F_f(1) = \int f(x) d\mu = \xi(f) \end{aligned}$$

for all $f \in L^2(\mathcal{B}, \mu)$ and $\xi \circ U_{\kappa}^* = \xi$ is proven, i.e., ξ is also invariant under U_{κ}^* . Theorem 2.2 and theorem 2.4 yield

Corollary 5.1. *If $U_\kappa(L^2(\mathcal{C}, \mu)) \subset L^2(\mathcal{C}, \mu)$, then the adjoint $U_\kappa^* : L^2(\mathcal{B}, \mu) \rightarrow L^2(\mathcal{B}, \mu)$ has a factor $\widehat{U_\kappa^*}$ with respect to the sub- σ -algebra \mathcal{C} .*

In the sequel we want to find sub- σ -algebras \mathcal{C} of \mathcal{B} which are invariant under the linear map U_κ induced by the Markov kernel κ . For point transformations $\varphi : X \rightarrow X$ we proved in corollary 4.3 of the previous section that sub- σ -algebras which are induced by eigenfunctions of the corresponding operator U_φ are invariant under U_φ . An analogous result does not hold true for eigenfunctions of the operator U_κ . Example 3.1 provides an even finite dimensional counterexample. The vector $(-1, 1, 0, 0)^T$ is an eigenvector of the Markov kernel T with eigenvalue 0. The σ -algebra induced by this eigenvector corresponds to its level sets and would suggest a possible aggregation of the third and the fourth state. But missing two further eigenvectors having the same level set structure, it follows from corollary 3.3 that such an aggregation is not possible at all, and therefore, the σ -algebra on the 4-state-space induced by $(-1, 1, 0, 0)^T$ is not invariant under the Markov kernel T . The crucial point in the derivation of corollary 4.3 was the fact that the preimages of point transformation φ maps sets onto sets. This is equivalent to say that the induced operator U_φ maps characteristic functions onto characteristic ones. Even more, the induced operator $U_\varphi : L^\infty(\mathcal{B}, \mu) \rightarrow L^\infty(\mathcal{B}, \mu)$ is not only linear but also a $*$ -homomorphism of the von Neumann Algebra $L^\infty(\mathcal{B}, \mu)$ (see [7] for further explanations), which turns out to be a sufficient conditions for an operator to map characteristic functions onto characteristic ones.

Following this line of thought we want to reduce the problem how to find sub- σ -algebras \mathcal{C} of \mathcal{B} which are invariant under the operator U_κ to the problem to find invariant subspaces V of $L^\infty(\mathcal{B}, \mu)$ such that the restriction of the operator U_κ on V is a $*$ -homomorphism. The work of S. Albeverio and R. Høegh-Krohn [1] addresses this problem for a particular class of Markov kernels.

Definition 5.2. We call a Markov kernel κ *ergodic* if the state ξ induced by the probability measure μ is the only invariant state of the adjoint U_κ^* .

Remark 5.1. Let κ be ergodic, and $\kappa(f) = f$ for an $f \in L^\infty(\mathcal{B}, \mu)$. This yields

$$\xi(fU_\kappa^*g) = \xi(U_\kappa(f)g) = \xi(fg)$$

for all $g \in L^2(\mathcal{B}, \mu)$ which proves that the state $\xi(f \cdot)$ is also invariant under U_κ^* . Ergodicity forces f to be proportional to the identity. Hence, definition 5.2 is analogues to the one for Markov kernels in the finite dimensional setting when κ is described by a stochastic matrix. In the finite case ergodicity is equivalent to the fact that there is only a single eigenfunction with eigenvalue 1. See theorem 1.19 in [18].

The following proof is based on the work of S. Albeverio and R. Høegh-Krohn [1]. Let $u_\alpha \in L^\infty(\mathcal{B}, \mu)$ be a normalized *root vector* of U_κ , i.e., $U_\kappa u_\alpha = \alpha u_\alpha$, $\alpha \in S^1$, where S^1 denotes the unit circle, and $\|u_\alpha\| = 1$. We have $U_\kappa \overline{u_\alpha} = \overline{U_\kappa u_\alpha} = \overline{\alpha u_\alpha} = \overline{\alpha} \overline{u_\alpha}$. We define a bounded functional on $L^\infty(\mathcal{B}, \mu)$ via $f \mapsto \xi(u_\alpha f \overline{u_\alpha})$ and denote it with τ . We have $\tau(1) = \xi(u_\alpha \overline{u_\alpha}) = \|u_\alpha\|^2 = 1$, thus τ is also a state on $L^\infty(\mathcal{B}, \mu)$. Furthermore, we have

$$\begin{aligned} \tau(U_\kappa^* f) &= \xi(u_\alpha U_\kappa^*(f) \overline{u_\alpha}) = \xi(u_\alpha \overline{u_\alpha} U_\kappa^* f) = \xi(U_\kappa(u_\alpha \overline{u_\alpha}) f) \\ &\geq \xi(U_\kappa(u_\alpha) U_\kappa(\overline{u_\alpha}) f) = \xi(u_\alpha f \overline{u_\alpha}) = \tau(f), \end{aligned}$$

but $\tau(U_\kappa^* f) \leq \tau(f)$ follows from the fact that $\|U_\kappa^*\| = \|U_\kappa\| \leq 1$. Hence $\tau \circ U_\kappa^* = \tau$ and we have $\xi = \tau$. From this and the fact that ξ is faithful one obtains $u_\alpha \overline{u_\alpha} = 1$ a.e. We define the positive semi definite form

$$\mu(f, f) = \xi(U_\kappa(\overline{f}f) - \overline{U_\kappa f} U_\kappa f).$$

We have $\mu(\overline{u_\alpha}, \overline{u_\alpha}) = 0$. From Schwarz inequality we derive

$$\mu(\overline{u_\alpha}, f) = 0$$

for all $f \in L^\infty(\mathcal{B}, \mu)$ which yields in combination with the faithfulness of the state ξ that for all $f \in L^\infty(\mathcal{B}, \mu)$

$$U_\kappa(u_\alpha f) - \alpha u_\alpha U_\kappa f = 0. \tag{5.2}$$

Let us denote with $\Gamma(\kappa) \subset S^1$ the set of all unitary roots of U_κ , i.e., $\alpha \in \Gamma(\kappa)$ iff there is a $u_\alpha \in L^\infty(\mathcal{B}, \mu)$ different from zero such that $U_\kappa u_\alpha = \alpha u_\alpha$. We have already seen that for every $\alpha \in \Gamma(\kappa)$ also $\bar{\alpha} \in \Gamma(\kappa)$. Let $\beta \in \Gamma(\kappa)$ and $u_\beta \in L^\infty(\mathcal{B}, \mu)$ an eigenfunction. From Eq. (5.2) we obtain

$$U_\kappa(u_\alpha u_\beta) = \alpha u_\alpha U_\kappa u_\beta = \alpha \beta u_\alpha u_\beta = U_\kappa u_\alpha U_\kappa u_\beta \quad (5.3)$$

which proves that also $\alpha\beta \in \Gamma(\kappa)$. Hence, $\Gamma(\kappa)$ is a subgroup of S^1 .

Definition 5.3. Let be $G \subset \Gamma(\kappa)$. The G -root- σ -algebra $\sigma(G)$ is generated by all sets

$$u_\alpha^{-1}(U) \quad \text{for all } U \subset \mathbb{C} \text{ Borel measurable, } \alpha \in G$$

Theorem 5.2. Let κ be ergodic and $G \subset \Gamma(\kappa)$. The adjoint of the Markov kernel $U_\kappa^* : L^2(\mathcal{B}, \mu) \rightarrow L^2(\mathcal{B}, \mu)$ has a factor with respect to the sub- σ -algebra $\sigma(G) \subset \mathcal{B}$.

Proof. From the spectral theorem in operator theory, see [7], one obtains that $L^\infty(\sigma(G), \mu)$ is a von Neumann algebra generated by the root vectors u_α with $\alpha \in G$. From Eq. (5.3) follows that the restriction of U_κ on $L^\infty(\sigma(G), \mu)$ is a $*$ -automorphism. In particular, we have $U_\kappa(L^\infty(\sigma(G), \mu)) = L^\infty(\sigma(G), \mu)$. Since $L^\infty(\sigma(G), \mu)$ is dense in $L^2(\sigma(G), \mu)$ and U_κ continuous with respect to the L^2 -norm, we obtain $U_\kappa(L^2(\sigma(G), \mu)) \subset L^2(\sigma(G), \mu)$. The assumption follows from the theorems 2.2 and 2.4. \square

6 Conclusions

We developed a systematic method to identify closed sub-dynamics of a given time-discrete dynamical system. These sub-systems lead to possible aggregations in the sense of Fig. (1) providing a reduced and self-contained description of the initial dynamical system. We worked out in detail that our approach is a proper generalization of aggregation methods well known for time-discrete dynamical systems on finite state spaces to the case of infinite state spaces. Furthermore, on two examples, point transformations and Markov processes, we demonstrated that the general framework does not only provide conceptual insights but also concrete solutions to the problem how to find possible aggregations for those time-discrete dynamical systems. Finally, the concept of an adjoint operator for Banach spaces developed in this paper is new and from a pure mathematical point of view interesting in its own right.

Ongoing research will be devoted to get rid of two restrictions: first, we want to deal with unbounded operators T whose adjoint are not defined in general; second, we want to consider even time-continuous dynamical systems. The first problem is of main importance to find possible aggregations of differential operators on manifolds, like the Liouville operator [15]. Those operators can be described as unbounded operators on a Hilbert space. The second problem should allow for a solution within our frame due to the Stone von Neumann theorem well know in quantum field theory in order to solve the Schrödinger equation. Due to this theorem every strongly continuous unitary one-parameter group is induced by a single self-adjoint operator. Hence, the problem studying the time continuous system derived from the one-parameter group is reduced to an investigation of the self-adjoint generator, where our techniques apply again.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 267087 and n° 318723.

References

- [1] S. Albeverio and R. Høegh-Krohn, *Frobenius theory for positive maps of von neumann algebras*, Commun. math. Phys. **64** (1978), 83–94.

- [2] P. Buchholz, *Hierarchical markovian models: symmetries and reduction*, Performance Evaluation **22** (1995), no. 1, 93–110.
- [3] C. K. Burke and M. Rosenblatt, *A markovian function of a markov chain*, Annals of Statistics **29** (1958), 1112–1122.
- [4] L. Capra, C. Dutheillet, G. Franceschinis, and J. Ilić, *On the use of partial symmetries for lumping markov chains*, ACM SIGMETRICS Performance Evaluation Review **28** (2001), no. 4, 33–35.
- [5] G. Cobb and Y. Chen, *An application of markov chain monte carlo to community ecology*, The American Mathematical Monthly **110** (2003), no. 4, 265–288.
- [6] P. Deuffhard, W. Huisinga, A. Fischer, and Ch. Schütte, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains*, Linear Algebra and its Applications **315** (2000), no. 1–3, 39–59.
- [7] J. Dixmier, *Von neumann algebras*, North-Holland Publishing Company, 1981.
- [8] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, 1981.
- [9] Olof Görnerup and Martin Nilsson Jacobi, *A dual digenvektor condition for strong lumpability of markov chains*, SFI WORKING PAPER (2008), 1–7.
- [10] Olof Görnerup and Martin Nilsson Jacobi, *A method for finding aggregated representations of linear dynamical systems*, Adv.Complex Syst. **13** (2010), no. 2, 199–215.
- [11] ———, *A model-independent approach to infer hierarchical codon substitution dynamics.*, BMC Bioinformatics **11** (2010), 201 (eng).
- [12] N. Israeli and N Goldenfeld, *Coarse-graining of cellular automata, emergence, and the predictability of complex systems*, Physical Review E **73** (2006), 026203.
- [13] J. L. Snell J. G. Kemeny, *Finite markov chains*, Springer-Verlag, 1976.
- [14] M. Nilsson Jacobi and O. Görnerup, *A spectral method for aggregating variables in linear dynamical systems with application to cellular automata renormalization*, Advances in Complex Systems **12** (2009), no. 2, 131–155.
- [15] Martin Nilsson Jacobi, *Hierarchical dynamics*, Encyclopedia of Complexity and Systems Science (Robert A. Meyers, ed.), Springer New York, 2009, pp. 4588–4608.
- [16] O. Pfante, E. Olbrich, N. Bertschinger, N. Ay, and J. Jost, *Comparison between different methods of level identification*, to appear (2013).
- [17] J.E. Rowe, M. Vose, and A. Wright, *State aggregation and population dynamics in linear systems*, Artificial Life **11** (2005), no. 4, 473–492.
- [18] P. Walters, *An introduction to ergodic theory*, Springer, 1982.