GMD Report
62

GMD –
Forschungszentrum
Informationstechnik
GmbH

Herbert Jaeger

Characterizing Distributions of Stochastic Processes by Linear Operators

September 1999

GMD –
Forschungszentrum Informationstechnik GmbH
Schloß Birlinghoven
D-53754 Sankt Augustin
Germany
Telefon +49 -2241 -14 -0
Telefax +49 -2241 -14 -2618
http://www.gmd.de

Anschrift des Verfassers/Address of the author:
Dr. Herbert Jaeger
Institut für Autonome intelligente Systeme
GMD – Forschungszentrum Informationstechnik GmbH
D-53754 Sankt Augustin
E-mail: Herbert.Jaeger@gmd.de

ISSN 1435-2702

Characterizing distributions of stochastic processes by linear operators

Herbert Jaeger
GMD – German National Research Center
for Information Technology

September 27, 1999

Abstract. This report describes how the finite-dimensional marginal distributions of a stochastic process $(\Omega, \mathfrak{A}, P, (X_t)_{t\in T})$ with values in an arbitrary measurable space $(E, \mathfrak{B})$, where $T = \mathbb{N}$ or $T = \mathbb{R}_{\geq 0}$, can be calculated with linear operators. In the discrete-time, discrete-observation case, we construct a family of linear operators $(\tau_a)_{a\in E}$ on a suitable real vector space V, such that any probability $P(X_0 = a_0, \ldots, X_{n-1} = a_{n-1})$ can be obtained by evaluating a product of matrices and vectors of the form $\sigma\tau_{a_{n-1}} \cdots \tau_{a_0} v_0$, where $v_0 \in V$ is a suitable starting vector and σ is the vector $(1, \ldots, 1)$. Conversely, necessary and sufficient conditions are given for algebraic structures $(V, (\tau_a)_{a\in E}, v_0)$ to describe a distribution of a process. Thus, a 1–1 correspondence between certain algebraic structures and distributions of discrete processes is found. These results are generalized to continuous time and arbitrary measurable spaces. All in all, one obtains a novel access to stochastic processes, by showing a way to model their distributions purely by means from linear algebra.

key words: stochastic processes, distributions, observable operator models

Zusammenfassung. In diesem Report wird beschrieben, wie die endlichdimensionalen Randverteilungen eines stochastischen Prozesses $(\Omega, \mathfrak{A}, P, (X_t)_{t\in T})$ mit Werten in einem beliebigen Meßraum $(E, \mathfrak{B})$, wobei $T = \mathbb{N}$ oder $T = \mathbb{R}_{\geq 0}$, mit linearen Operatoren berechnet werden können. Im Falle diskreter Zeit und endlichen Meßraumes konstruieren wir eine Familie linearer Operatoren $(\tau_a)_{a\in E}$ auf einem geeigneten reellen Vektorraum V, so dass Wahrscheinlichkeiten der Art $P(X_0 = a_0, \ldots, X_{n-1} = a_{n-1})$ durch die Auswertung eines Matrizen/Vektorproduktes der Form $\sigma\tau_{a_{n-1}} \cdots \tau_{a_0} v_0$ erhalten werden, wobei $v_0 \in V$ ein geeigneter Startvektor und σ der Vektor $(1, \ldots, 1)$ ist. Umgekehrt werden hinreichende und notwendige Bedingungen für eine algebraische Struktur $(V, (\tau_a)_{a\in E}, v_0)$ angegeben, dass diese in der erwähnten Weise die Verteilung eines Prozesses beschreibt. So wird eine $1-1$ Beziehung zwischen bestimmten algebraischen Strukturen und Verteilungen diskreter Prozesse aufgewiesen. Diese Ergebnisse werden sodann auf kontinuierliche Zeit und beliebige Messräume verallgemeinert. Alles in allem erhält man einen neuen Zugang zu stochastischen Prozessen, indem deren Verteilungen mit Mitteln der linearen Algebra modelliert werden.

Stichwörter: stochastische Prozesse, Verteilungen, observable operator models

1 Introduction.

In this article we describe how the distribution of a stochastic process can be characterized through certain linear operators, called the *observable operators* of the process.

The paper deals with distributions of stochastic processes $(\Omega, \mathfrak{A}, P, (X_t)_{t\in T})$, where $T = \mathbb{N}$ or $T = \mathbb{R}_{\geq 0}$, and the random variables take values in an arbitrary measurable space $(E, \mathfrak{B})$. For a first impression, we state here a main result and the resulting definition for the special case of discrete-time, finite-valued processes.

Proposition 1 *Let $(\Omega, \mathfrak{A}, P, (X_n)_{n\in\mathbb{N}})$ be a discrete-time, stochastic process with values in a finite set $E = \{a_1, \ldots, a_k\}$. Then there exists a real vector space $\mathfrak{G}$, a basis $(\mathfrak{e}_j)_{j\in J}$ of $\mathfrak{G}$, a vector $\mathfrak{g}_\varepsilon \in \mathfrak{G}$, and a family of linear operators $(\mathfrak{t}_a)_{a\in E}$ indexed by the possible values of the process, such that all probabilities $P(X_0 = a_{i_0}, \ldots, X_m = a_{i_m})$ of finite-length initial realizations of the process can be computed in the following way. Apply the operators $\mathfrak{t}_{a_{i_0}}, \ldots, \mathfrak{t}_{a_{i_m}}$ consecutively to $\mathfrak{g}_\varepsilon$, and let $\sum_{i=1,\ldots,\nu} \alpha_i \mathfrak{e}_{j_i} = \mathfrak{t}_{a_{i_m}} \circ \cdots \circ \mathfrak{t}_{a_{i_0}} \mathfrak{g}_\varepsilon$ be the linear combination of the resulting vector from basis vectors. Then obtain the desired probability through*

$$P(X_0 = a_{i_0}, \ldots, X_m = a_{i_m}) = \sum_{i=1,\ldots,\nu} \alpha_i. \tag{1}$$

Observe that the distribution of $(\Omega, \mathfrak{A}, P, (X_n)_{n\in\mathbb{N}})$ is completely determined by the probabilities of finite-length initial realizations. This motivates the following definition:

Definition 1 *A structure $\mathcal{A} = (\mathfrak{G}, (\mathfrak{e}_j)_{j\in J}, (\mathfrak{t}_a)_{a\in E}, \mathfrak{g}_\varepsilon)$, from which the finite-length initial realizations of a stochastic process $(\Omega, \mathfrak{A}, P, (X_n)_{n\in\mathbb{N}})$ can be computed according to (1), is an* observable operator model *(OOM) of the distribution of the process.*

The name, *observable operators*, is motivated by the fact that possible observations (i.e., symbols a_i) correspond 1–1 with the operators $\mathfrak{t}_{a_i}$ in the sense that the probabilities of *observations* $a_{i_0} \ldots a_{i_m}$ are modeled by chains of *operators* $\mathfrak{t}_{a_{i_m}} \circ \cdots \circ \mathfrak{t}_{a_{i_0}}$. OOMs allow to characterise and analyse distributions of stochastic processes purely by means of linear algebra.

An important special case is obtained when the vector space $\mathfrak{G}$ has finite dimension. The observable operators can then be represented by matrices, and methods of numerical linear algebra can be applied to various practical tasks connected with stochastic systems, like system identification or prediction. Specifically, linear algebra methods lead to a fast, constructive algorithm for learning observable operator models from empirical data. This is remarkable since finite-dimensional OOMs subsume hidden Markov models, for which only iterative learning algorithms are known. The topics of matrix representations, prediction, learning algorithm, and relation to hidden Markov models are detailed out in [17]; a generalization to discrete input-output systems is given in [16].

Two lines of research in probability theory are related to observable operator models.

Firstly, the idea that observations $a_{i_0} \dots a_{i_m}$ correspond to a sequence of operators is also constitutional for the theory of *random systems with complete connections* [13]. Besides sharing this idea, the two approaches have not much in common. Specifically, the operators in random systems with complete connections need not be linear, and distributions are not characterized through the operators.

Secondly, OOM theory is strongly related to a strand of research [10] [4] [5] [6] [7] [8] [9], which had been triggered by the question when two hidden Markov models are equivalent, i.e. describe the same distribution [3]. In the course of these investigations, a number of ways was described how to compute probabilities by sequences of matrix multiplications. These computations are analogous to (1), although the obtained matrices were in general different (and higher-dimensional) than the matrices obtained in finite-dimensional OOMs [17]. These matrix-based techniques culminated in a theorem that gave necessary and sufficient conditions for equivalence of hidden Markov models (HMMs), thereby answering the initial question [15]. A different approach was taken by [12]. Instead of using matrix representations, Heller developed *stochastic modules*. They provide a representation of general discrete-time, discrete-valued processes. Stochastic modules are defined within a framework of category theory and module theory. The approach was forgotten until in a recent dissertation thesis [14], stochastic modules were connected to the more common matrix representations. This dissertation is also a comprehensive presentation of the entire strand of research.

Both stochastic modules and OOMs provide algebraic models of all discrete-time, discrete-valued processes. Two facts justify the introduction of OOMs. First, they are mathematically more elementary (linear algebra vs. category theoretic module theory in stochastic modules), and thus more accessible. Second, while previous work was confined to discrete-time, discrete valued systems, OOM theory can straightforwardly be extended to continuous-time, arbitrary-valued processes – which is done in this article.

2 Construction of discrete-time, discrete-value OOMs.

For didactic reasons, we treat the special case of discrete-time, finite-value processes seperately from the general case. This section contains a proof of Proposition 1, introduces the basic constructions of OOM theory, and presents a number of other theorems which are central to the theory.

Let $(\Omega, \mathfrak{A}, P, (X_n)_{n\in\mathbb{N}})$ be a stochastic process with values in a finite set $E = \{a_1, \ldots, a_k\}$. For simplicity, in the remainder of this section we will call such processes *discrete processes*, and their distributions, *discrete distributions*. We construct a vector space $\mathfrak{G}$, a basis $(\mathfrak{e}_j)_{j\in J}$, a vector $\mathfrak{g}_\varepsilon \in \mathfrak{G}$, and a family of operators $(\mathfrak{t}_a)_{a\in E}$ such that the statement of theorem 1 becomes true.

Let $E^n = \{a_0 \ldots a_{n-1} \mid a_0, \ldots, a_{n-1} \in E\}$ denote the set of all initial observations of length n from elements of E $(n \geq 1)$. Formally, $a_0 \ldots a_{n-1}$ is a *word* over the alphabet E. It is a general notational characteristic of the OOM approach that observations $X_0 = a_{i_0}, \ldots, X_{n-1} = a_{i_{n-1}}$ are denoted by words $a_{i_0} \ldots a_{i_{n-1}}$. For $n = 0$, we put $E^0 = \{\varepsilon\}$: ε – the *empty word* – denotes the absence of any observation. Let $E^* = \bigcup_{n\geq 0} E^n$ be the set of all words.

We use the shorthand $\bar{b}$ to denote any element of E^*. Furthermore, for $\bar{a} = a_0 \ldots a_{n-1}, \bar{b} = b_0 \ldots b_{m-1} \in E^*$ we introduce the shorthand $P(\bar{a} \mid \bar{b})$ to denote $P(X_m = a_0, \ldots, X_{m+n-1} = a_{n-1} \mid X_0 = b_0, \ldots, X_{m-1} = b_{m-1})$ in the case $m, n \geq 1$, to denote $P(X_0 = a_0, \ldots, X_{n-1} = a_{n-1})$ in the case $n \geq 1, \bar{b} = \varepsilon$, and to denote 1 in the case $\bar{a} = \varepsilon$. Furthermore, the shorthand $P(\bar{a})$ denotes $P(X_0 = a_0, \ldots, X_{n-1} = a_{n-1})$ in the case $n \geq 1$, and it denotes 1 in the case $\bar{a} = \varepsilon$.

For every $\bar{b} \in E^*$ we define a numerical function $\mathfrak{g}_{\bar{b}} : E^* \to \mathbb{R}$ by putting

$$\mathfrak{g}_{\bar{b}}(\bar{a}) = \begin{cases} P(\bar{a} \,|\, \bar{b}), & \text{if } P(\bar{b}) > 0 \\ 0, & \text{if } P(\bar{b}) = 0 \end{cases} \tag{2}$$

for all $\bar{a} \in E^*$.

In the remainder of this article, we will tacitly omit case distinctions such as in 2, which account for zero probabilities of conditioning events, because they are obvious.

The function $\mathfrak{g}_{\bar{b}}$ can be understood as a prediction function: $\mathfrak{g}_{\bar{b}}(\bar{a})$ gives the probability that an initially obtained observation $\bar{b}$ will be continued by $\bar{a}$. If no prior observation is available, (i.e., $\bar{b} = \varepsilon$), then $\mathfrak{g}_{\bar{b}}(\bar{a}) = \mathfrak{g}_\varepsilon(\bar{a})$ gives the unconditioned probability to observe $\bar{a}$. Thus, each $\mathfrak{g}_{\bar{b}}$ specifies a conditional distribution (case $P(\bar{b}) > 0$) or is the null vector (case $P(\bar{b}) = 0$). Note that the distribution specified by $\mathfrak{g}_\varepsilon$ is the very distribution of the process.

Let $\mathfrak{D}$ denote the set of all functions from E^* into the reals. $\mathfrak{D}$ canonically becomes a real vector space if one defines scalar multiplication and vector addition as follows: for $\mathfrak{d}_1, \mathfrak{d}_2 \in \mathfrak{D}$, $\alpha \in \mathbb{R}$, $\bar{a} \in E^*$ put $(\alpha\mathfrak{d}_1)(\bar{a}) = \alpha(\mathfrak{d}_1(\bar{a}))$ and $(\mathfrak{d}_1 + \mathfrak{d}_2)(\bar{a}) := \mathfrak{d}_1(\bar{a}) + \mathfrak{d}_2(\bar{a})$.

Now let $\mathfrak{G} = \langle\{\mathfrak{g}_{\bar{b}} \,|\, \bar{b} \in E^*\}\rangle_{\mathfrak{D}}$ be the linear subspace spanned in $\mathfrak{D}$ by all of the conditional distributions introduced above. Furthermore, choose $E_0 \subseteq E^*$ such that $(\mathfrak{e}_j)_{j\in J} = \{\mathfrak{g}_{\bar{c}} \,|\, \bar{c} \in E_0\} \subset \{\mathfrak{g}_{\bar{b}} \,|\, \bar{b} \in E^*\}$ is a basis of $\mathfrak{G}$ (this implies that every $\bar{c} \in E_0$ has a positive probability).

Thus we have constructed the vector space $\mathfrak{G}$, the basis $(\mathfrak{e}_j)_{j\in J}$, and the special vector $\mathfrak{g}_\varepsilon$. It remains to specify the family of observable operators $(\mathfrak{t}_a)_{a\in E}$.

In order to specify a linear operator on $\mathfrak{G}$, it suffices to specify the values the operator takes on the basis vectors. Define, for every $a \in E$, a linear operator $\mathfrak{t}_a : \mathfrak{G} \to \mathfrak{G}$ by putting

$$\mathfrak{t}_a \mathfrak{g}_{\bar{c}} = P(a \,|\, \bar{c})\; \mathfrak{g}_{\bar{c}a} \tag{3}$$

for all $\bar{c} \in E_0$ ($\bar{c}a$ denotes the sequence obtained by appending a to $\bar{c}$). It turns out that (3) carries over from basis elements $\{\mathfrak{g}_{\bar{c}} \,|\, \bar{c} \in E_0\}$ to all predictors $\mathfrak{g}_{\bar{b}}$ (where $\bar{b} \in E^*$):

Proposition 2 *The linear operators $\mathfrak{t}_a$ defined through (3) satisfy the condition*

$$\mathfrak{t}_a \mathfrak{g}_{\bar{b}} = P(a \,|\, \bar{b})\; \mathfrak{g}_{\bar{b}a} \tag{4}$$

for all $\bar{b} \in E^*$, which satisfy $P(\bar{b}) > 0$.

Proof. Let $\mathfrak{g}_{\bar{b}} = \sum_{i=1}^{k} \alpha_i \mathfrak{g}_{\bar{c}_i}$ be the linear combination of $\mathfrak{g}_{\bar{b}}$ from k basis elements taken from $\{\mathfrak{g}_{\bar{c}} \,|\, \bar{c} \in E_0\}$. Let $\bar{d} \in E^*$. Then obtain (4) through the following calculation:

$$\begin{aligned}
(\mathfrak{t}_a \mathfrak{g}_{\bar{b}})(\bar{d}) \quad &= \\
= \quad (\mathfrak{t}_a(\sum_{i=1}^{k} \alpha_i \mathfrak{g}_{\bar{c}_i}))(\bar{d}) \quad &= \quad (\sum \alpha_i P(a \,|\, \bar{c}_i)\; \mathfrak{g}_{\bar{c}_i a})(\bar{d}) \\
= \quad \sum \alpha_i P(a \,|\, \bar{c}_i) P(\bar{d} \,|\, \bar{c}_i a) \quad &= \quad \sum \alpha_i \frac{P(\bar{c}_i) P(a\bar{d} \,|\, \bar{c}_i)}{P(\bar{c}_i)} \\
= \quad \mathfrak{g}_{\bar{b}}(a\bar{d}) \quad = \quad P(a\bar{d} \,|\, \bar{b}) \quad &= \quad P(a \,|\, \bar{b})\; P(\bar{d} \,|\, \bar{b}a) \\
= \quad P(a \,|\, \bar{b})\; \mathfrak{g}_{\bar{b}a}(\bar{d}). \quad \square &
\end{aligned}$$

We now restate and prove the claim made in the introduction.

Proposition 3 *Let $(\Omega, \mathfrak{A}, P, (X_n)_{n\in\mathbb{N}})$ be a stochastic process with values in a finite set $E = \{a_1, \ldots, a_k\}$, and let $\mathcal{A} = (\mathfrak{G}, (\mathfrak{g}_{\bar{b}_j})_{j\in J}, (\mathfrak{t}_a)_{a\in E}, \mathfrak{g}_\varepsilon)$ be a structure derived from the process, as described above. Then the distribution of the process can be calculated through $\mathcal{A}$ according to (1).*

Proof. Let $\sum_{i=1,\ldots,k} \alpha_i \mathfrak{g}_{\bar{c}_i} = \mathfrak{t}_{a_{i_m}} \ldots \mathfrak{t}_{a_{i_0}} \mathfrak{g}_\varepsilon$ be the linear combination of $\mathfrak{t}_{a_{i_m}} \ldots \mathfrak{t}_{a_{i_0}} \mathfrak{g}_\varepsilon$ from basis vectors. From an iterated application of (4) it follows that $\mathfrak{t}_{a_{i_m}} \ldots \mathfrak{t}_{a_{i_0}} \mathfrak{g}_\varepsilon = P(a_{i_0} \ldots a_{i_m}) \;\; \mathfrak{g}_{a_{i_0} \ldots a_{i_m}}$. Using this fact, we conclude

$$\begin{aligned}
&P(a_{i_0} \ldots a_{i_m}) = \\
&\quad = \;\; P(a_{i_0} \ldots a_{i_m}) \sum_{b\in E} \mathfrak{g}_{a_{i_0} \ldots a_{i_m}}(b) \quad = \quad \sum_{b\in E} \mathfrak{t}_{a_{i_0}} \cdots \mathfrak{t}_{a_{i_m}} \mathfrak{g}_\varepsilon(b) \\
&\quad = \;\; \sum_{i=1,\ldots,k} \alpha_i \sum_{b\in E} \mathfrak{g}_{\bar{c}_i}(b) \quad = \quad \sum_{i=1,\ldots,k} \alpha_i. \quad \square
\end{aligned}$$

The vector space $\mathfrak{G}$ is uniquely determined by the distribution. Its dimension is therefore a characteristic of the distribution:

Definition 2 *The dimension of the vector space $\mathfrak{G}$ is the* dimension *of the distribution.*

Remark 1 *Intuitively, vectors $\mathfrak{g}_{\bar{a}}$ and operators $\mathfrak{t}_a$ are best understood as modeling an observer's* knowledge about the state *of an observed process. In physics, control engineering, and the system sciences, a state s_t at time t of a temporal system is defined as that inside the system which determines the system's future (formal treatment in [18]). What an observer knows about a system's state, then, is tantamount to what he or she knows about the system's future. We have construed the vectors $\mathfrak{g}_{\bar{a}}$ as stochastic prediction functions. $\mathfrak{g}_{\bar{a}}$ models what can be said statistically about the system's future, given prior observations $\bar{a}$. Thus, $\mathfrak{g}_{\bar{a}}$ reflects an observer's knowledge about the system. Specifically, it would be inappropriate to interpret $\mathfrak{g}_{\bar{a}}$ as a state of the system itself. Likewise, the operator $\mathfrak{t}_a$ describes the change of knowledge about a process due to an incoming observation of a. These issues become especially clear when one considers OOMs of hidden Markov models, where both proper system states (the hidden states of the underlying Markov process) and "knowledge states" (the vectors of the OOM) are formally defined. OOMs of hidden Markov models are extensively treated in [17].*

So far, we have seen that every discrete distribution can be modeled by an OOM $\mathcal{A}$. We now consider the converse: given a structure $\mathcal{A} = (V, (e_j)_{j\in J}, (\tau_a)_{a\in E}, v_0)$, (where V is a real vector space with basis $(e_j)_{j\in J}$, $(\tau_a)_{a\in E}$ are linear operators on V, $v_0 \in V$), under which conditions does there exist a discrete distribution, of which $\mathcal{A}$ is an observable operator model?

We introduce the following notation. Let V be a real vector space with basis $(e_j)_{j\in J}$. Let $\sigma_{(e_j)_{j\in J}} : V \to \mathbb{R}$ be the function which assigns to every vector $w \in V$ the sum of coefficients from its basis vector combination, i.e.

$$\sigma_{(e_j)_{j\in J}}(w) = \sum_{i=1,\dots,k} \alpha_i, \tag{5}$$

where $w = \sum_{i=1,\dots,k} \alpha_i e_{j_i}$. Note that σ is linear. We drop the subscript $(e_j)_{j\in J}$ from σ when the reference to the basis is clear. Furthermore, for a word $a_0 \dots a_m = \bar{a} \in E^*$ we introduce the shorthand $\tau_{\bar{a}}$ to denote the concatenation $\tau_{a_{i_m}} \circ \cdots \circ \tau_{a_{i_0}}$ in the case $m \geq 0$, and to denote the identity mapping **id** in the case $\bar{a} = \varepsilon$.

Proposition 4 *Let V be a real vector space with basis $(e_j)_{j\in J}$, $(\tau_a)_{a\in E}$ a familiy of linear operators on V which is indexed by a finite set E, and $v_0 \in V$. Assume furthermore that V is spanned by the vectors $\{\tau_{\bar{a}} v_0 \,|\, \bar{a} \in E^*\}$. Define a numerical function $P : E^* \to \mathbb{R}$ by putting $P(\bar{a}) = \sigma\tau_{\bar{a}} v_0$, and $P(\varepsilon) = 1$. Let $\mu := \sum_{a\in E} \tau_a$. Then P can be extended to the distribution of a discrete stochastic process, if and only if the following three conditions hold:*

1. *$\sigma v_0 = 1$,*
2. *$\sigma\mu e_j = \sigma e_j$ for all basis vectors e_j,*
3. *for all sequences $\bar{a} \in E^*$ it holds that $\sigma\tau_{\bar{a}} v_0 \geq 0$.*

Proof. $\Longleftarrow$: Recall that a numerical function $P : E^* \to \mathbb{R}$ can be (uniquely) extended to the distribution of a discrete stochastic process, if the following two conditions are met:

A. For all $n \geq 0$, P is a probability measure on the observations of length n. I.e., (i) $P(a_{i_0} \dots a_{i_{n-1}}) \geq 0$, and (ii) $\sum_{a_{i_0} \cdots a_{i_{n-1}} \in E^n} P(a_{i_0} \dots a_{i_{n-1}}) = 1$.

B. P is consistent, i.e., $P(a_{i_0} \dots a_{i_{n-1}}) = \sum_{b\in E} P(a_{i_0} \dots a_{i_{n-1}} b)$.

A.(i) is a consequence of condition *3*. A.(ii) and B can easily be derived from conditions *1* and *2*, if one observes that condition *2* implies $\sigma\mu w = \sigma w$ for all $w \in V$, and that $\sum_{a_{i_0} \cdots a_{i_{n-1}} \in E^n} \sigma\tau_{a_{i_{n-1}}} \cdots \tau_{a_{i_0}} v_0 = \sigma\mu \cdots \mu v_0$.

$\Longrightarrow$: Assume that P can be extended to a distribution. Conditions *1* and *3* follow trivially. For condition *2*, observe that $\sigma\mu\tau_{\bar{a}}v_0 = \sigma\sum_{b\in E}\tau_b\tau_{\bar{a}}v_0 = \sigma\tau_{\bar{a}}v_0$. Since V is spanned by the vectors $\{\tau_{\bar{a}}v_0 \,|\, \bar{a} \in E^*\}$, this implies that for all $w \in V$ it holds that $\sigma\mu w = \sigma w$. This subsumes condition *2* as a special case. □

Condition *3* is no help when it comes to constructing OOMs, because it gives no clues about how the matrices τ_a must be formed in order to satisfy this condition. This situation is much ameliorated by an equivalent version of condition *3*, which goes back to [12]. To spell out this version, we need some concepts from the theory of convex cones. We follow the notation of a standard textbook [2].

With a set $S \subseteq \mathbb{R}^n$ we associate the set S^G, the *set generated by* S, which consists of all finite nonnegative linear combinations of elements of S. A set $K \subseteq \mathbb{R}^n$ is defined to be a *convex cone* if $K = K^G$. A cone K is *pointed* if for every nonzero $v \in K$, the vector $-v$ is not in K.

Using these concepts, the following theorem gives a condition which can replace condition *3* in Proposition 4 by a practically more useful alternative.

Proposition 5 ([12]) *Let* $\mathcal{A} = (\mathbb{R}^m, (\tau_a)_{a\in E}, v_0)$ *be a structure consisting of linear maps* $(\tau_a)_{a\in E}$ *on* $\mathbb{R}^m$ *and a vector* $v_0 \in \mathbb{R}^m$*. Let* $\mu := \sum_{a\in E}\tau_a$*. Assume that the first two conditions from Proposition 4 hold. Then* $\mathcal{A}$ *is an OOM if and only if there exists a pointed convex cone* K *satisfying the following conditions:*

1. $\sigma v \geq 0$ *for all* $v \in K$,
2. $v_0 \in K$,
3. $\forall a \in E: \; \tau_a K \subseteq K$.

Notes on the proof, variants and extensions can be found in [17]. Proposition 5 can be used to build OOMs from scratch, starting with a cone K and constructing observable operators satisfying $\tau_a K \subseteq K$. Note, however, that the theorem provides no means to *decide*, for a given structure $\mathcal{A}$, whether $\mathcal{A}$ is a valid OOM, since the theorem is non-constructive w.r.t. K.

Remark 2 *A note on terminology. We write* $(\mathfrak{G}, (\mathfrak{e}_j)_{j\in J}, (\mathfrak{t}_a)_{a\in E}, \mathfrak{g}_\varepsilon)$ *for observable operator models whose vectors are construed as conditional probability functions, and whose operators are defined by (3). By contrast, in general we write* $(V, (e_j)_{j\in J}, (\tau_a)_{a\in E}, v_0)$ *for any linear algebra structure which specifies a projective measure, which in turn gives rise to the finite-dimensional* *marginal distributions of a stochastic process. The generic term for both* $(\mathfrak{G}, (\mathfrak{e}_j)_{j\in J}, (\mathfrak{t}_a)_{a\in E}, \mathfrak{g}_\varepsilon)$ *and* $(V, (e_j)_{j\in J}, (\tau_a)_{a\in E}, v_0)$ *is* observable operator model*; a specific term for the former kind of models is* canonical *OOMs (they were somewhat clumsily called* predictor-space *OOMs in [17]).*

Summing up, in this section we have described a correspondence between

a probabilistic concept: the distribution of a discrete stochastic process, and

a linear algebra concept: structures $(V, (e_j)_{j\in J}, (\tau_a)_{a\in E}, v_0)$, which satisfy the conditions from Proposition 4.

The correspondence becomes effective through the equation $P(\bar{a}) = \sigma\tau_{\bar{a}}v_0$. Up to liberty in the choice of the basis $(e_j)_{j\in J}$, the correspondence is 1 – 1. For finite-dimensional processes, all admissible changes of basis can be effectively constructed (detailed out in [17]).

3 OOMs of continuous-time, arbitrary-valued distributions.

In this section, we generalize the previous results to distributions of processes of the kind $(\Omega, \mathfrak{A}, P, (X_t)_{t\in\mathbb{R}_{\geq 0}})$, where the random variables X_t have values in an arbitrary measurable space $(E, \mathfrak{B})$.

A continuous-time distribution is specified through its finite-dimensional marginal distributions, i.e. the probabilities of the finite-dimensional cylinder sets in $\mathfrak{B}^{\mathbb{R}_{\geq 0}}$. These latter probabilities are in turn completely determined by the probabilities of particular kind of cylinder sets, which can be specified in the following way:

$$\{X_{t_0} \in A_0, \ldots, X_{t_{n-1}} \in A_{n-1} \,|\, n \geq 1, 0 \leq t_0 < \ldots < t_{n-1}, A_i \in \mathfrak{B}\}. \tag{6}$$

We shall call cylinder sets specified in the way of (6), *basic* cylinders. It is easy to see that the basic cylinders generate the cylinder sets (cf. [1] exercise in section 22). Since furthermore the set of basic cylinders is closed with respect to $\cap$, a standard theorem about the unique extension of finite measures (cf. [1], theorem 5.5) yields that the probabilities of basic cylinders uniquely determine the probabilities of finite-dimensional cylinder sets. Therefore, we can restrict ourselves to probabilities of basic cylinders if we wish to specify distributions of continuous-time processes.

In fact, the main work to be done in order to extend the discrete-time, discrete-value to the continuous-time, arbitrary-valued case, is to introduce a suitable notational convention, which makes it possible to write basic cylinders as words. Once that is done, the theorems and proofs of the previous section can be re-used without any change.

Definition 3 *(A word-based notation for basic cylinders and conditional probabilities.)*

Let

$$\begin{aligned} C &:= \{(A,r)\,|\,A\in\mathfrak{B}, r\in\mathbb{R}_{>0}\},\\ C^* &:= \{(A_0,r_0)\dots(A_{n-1},r_{n-1})\,|\,n\geq 1, A_i\in\mathfrak{B}, r_i\in\mathbb{R}_{>0}\}\cup\{\varepsilon\}\end{aligned}$$

denote the set of pairs of measurable events and positive real numbers, and the finite sequences thereof, respectively. We use symbols a, b, c for elements of C, $\bar{a}$ etc. for elements of C^*, and write $\bar{b}a$, $\bar{b}\bar{a}$ for concatenations in the obvious way (cf. figure 1). Use elements of C^* to denote basic cylinders, as follows:

1. *ε denotes the basic cylinder $\{X_0\in E\}$,*
2. *A word $\bar{a}\in C^*$ of the form $\bar{a}=(A_0,r_0)\dots(A_{n-1},r_{n-1})$, where $n\geq 1$, denotes the basic cylinder $\{X_0\in A_0,\ X_{r_0}\in A_1,\ X_{r_0+r_1}\in A_2,\ \dots,\ X_{r_0+\dots+r_{n-2}}\in A_{n-1}\}$.*

(Note that the first observation in basic cylinders according to 2. is made at time 0. A basic cylinder $\{X_{t_0}\in A_0,\ \dots,\ X_{t_{n-1}}\in A_{n-1}\}$, where the first observation is taken at $t_0>0$, can nevertheless be expressed in the way of 2. by re-writing $\{X_{t_0}\in A_0,\ \dots,\ X_{t_{n-1}}\in A_{n-1}\}$ equivalently as $\{X_0\in E,\ X_{t_0}\in A_0,\ \dots,\ X_{t_{n-1}}\in A_{n-1}\}$.)

We write $P(\bar{a})$ for the probability of the basic cylinder denoted by $\bar{a}$.

If $\bar{a}=(A_0,r_0)\dots(A_{n-1},r_{n-1})$ denotes $\{X_0\in A_0,\ \dots,\ X_{r_0+\dots+r_{n-2}}\in A_{n-1}\}$ and $\bar{b}=(B_0,s_0)\dots(B_{m-1},s_{m-1})$ denotes $\{X_0\in B_0,\ \dots,\ X_{s_0+\dots+s_{m-2}}\in B_{m-1}\}$, we write $P(\bar{a}\,|\,\bar{b})$ to denote the conditional probability $P(X_{s_0+\dots+s_{m-1}}\in A_0, X_{s_0+\dots+s_{m-1}+r_0}\in A_1,\ \dots,\ X_{s_0+\dots+s_{m-1}+r_0+\dots+r_{n-2}}\in A_{n-1}\,|\,X_0\in B_0,\ \dots,\ X_{s_0+\dots+s_{m-2}}\in B_{m-1})$.

Proposition 6 *Let $(\Omega,\mathfrak{A},P,(X_t)_{t\in\mathbb{R}_{\geq 0}})$ be a stochastic process, where the random variables X_t have values in an arbitrary measurable space $(E,\mathfrak{B})$. Then there exist a real vector space $\mathfrak{G}$, a basis $(\mathfrak{e}_j)_{j\in J}$ of $\mathfrak{G}$, a vector $\mathfrak{g}_\varepsilon\in\mathfrak{G}$,*

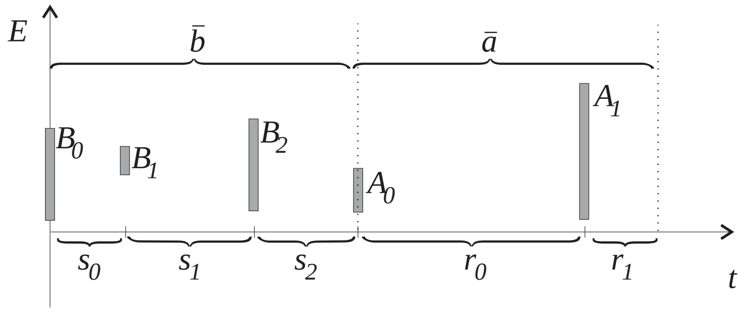

Figure 1: Example for definition 3. A word $\bar{b} = (B_0, s_0)(B_1, s_1)(B_2, s_2)$ concatenated with a word $\bar{a} = (A_0, r_0)(A_1, r_1)$ corresponds to the basic cylinder $\bar{b}\bar{a} = \{X_0 \in B_0, X_{s_0} \in B_1, X_{s_0+s_1} \in B_2, X_{s_0+s_1+s_2} \in A_0, X_{s_0+s_1+s_2+r_0} \in A_1\}$.

and a family of linear operators $(\mathfrak{t}_a)_{a\in C}$ indexed by measurable events and positive reals, such that the probability of the basic cylinder denoted by $\bar{a} = a_0 \cdots a_{n-1} = (A_0, r_0) \cdots (A_{n-1}, r_{n-1})$ can be computed in the following way:

$$\begin{aligned} &P(X_0 \in A_0, X_{r_0} \in A_1, \ldots, X_{r_0+\cdots+r_{n-2}} \in A_{n-1}) = \\ &= \sigma_{(\mathfrak{e}_j)_{j\in J}}(\mathfrak{t}_{(A_{n-1},r_{n-1})} \cdots \mathfrak{t}_{(A_1,r_1)}\mathfrak{t}_{(A_0,r_0)}\mathfrak{g}_\varepsilon). \end{aligned} \tag{7}$$

In shorthand notation, this becomes $P(\bar{a}) = \sigma \mathfrak{t}_{\bar{a}}\mathfrak{g}_\varepsilon$.

Proof. We construct $\mathfrak{G}$, $(\mathfrak{e}_j)_{j\in J}$, $\mathfrak{g}_\varepsilon$ and the observable operators in perfect analogy to the discrete case. The only difference lies in the semantics of words $\bar{a}$, which here denote basic cylinders.

For every $\bar{b} \in C^*$ (including $\bar{b} = \varepsilon$) we define a numerical function $\mathfrak{g}_{\bar{b}} : C^* \to \mathbb{R}$ by $\mathfrak{g}_{\bar{b}}(\bar{a}) = P(\bar{a} \mid \bar{b})$. Let $\mathfrak{D}$ denote the real vector space of all functions from C^* into the reals. Let $\mathfrak{G} = \langle\{\mathfrak{g}_{\bar{b}} \mid \bar{b} \in C^*\}\rangle_{\mathfrak{D}}$ be the linear subspace spanned in $\mathfrak{D}$ by all functions $\mathfrak{g}_{\bar{b}}$. Choose $C_0 \subseteq C^*$ such that $(\mathfrak{e}_j)_{j\in J} = \{\mathfrak{g}_{\bar{c}} | \bar{c} \in C_0\}$ is a basis of $\mathfrak{G}$.

For every $a \in C$, define a linear operator $\mathfrak{t}_a : \mathfrak{G} \to \mathfrak{G}$ by putting $\mathfrak{t}_a\mathfrak{g}_{\bar{c}} = P(a \mid \bar{c})\, \mathfrak{g}_{\bar{c}a}$ for all $\bar{c} \in C_0$. This defining equation carries over to all $\bar{c} \in C$ (re-use Proposition 2). Finally, re-use Proposition 3 to conclude the proof. □

Like in the discrete case, we call $\mathcal{A} = (\mathfrak{G}, (\mathfrak{e}_j)_{j\in J}, (\mathfrak{t}_a)_{a\in C}, \mathfrak{g}_\varepsilon)$ an observable operator model of the distribution of $(\Omega, \mathfrak{A}, P, (X_t)_{t\in\mathbb{R}_{\geq 0}})$, and the dimension of $\mathfrak{G}$ is taken as the dimension of the process.

The family $(\mathfrak{t}_a)_{a\in C}$ is internally structured by virtue of the following relationships:

Proposition 7 *Let* $(\mathfrak{G}, (\mathfrak{e}_j)_{j\in J}, (\mathfrak{t}_a)_{a\in C}, \mathfrak{g}_\varepsilon)$ *be an observable operator model of a process* $(\Omega, \mathfrak{A}, P, (X_t)_{t\in\mathbb{R}_{\geq 0}})$ *with values in measurable space* $(E, \mathfrak{B})$*. Then it holds that*

1. $\mathfrak{t}_{(\bigcup_{n=1}^{\infty} A_n, r)} = \sum_{n=1}^{\infty} \mathfrak{t}_{(A_n,r)}$ *for all sequences of pairwise disjoint* $A_i \in \mathfrak{B}$*,*
2. $\mathfrak{t}_{(A,r_1+r_2)} = \mathfrak{t}_{(E,r_2)} \circ \mathfrak{t}_{(A,r_1)}$ *for all* $r_1, r_2 > 0$*.*

Proof. *1.* We have to show that the operators $\mathfrak{t}_{(\bigcup_{n=1}^{\infty} A_n, r)}$ and $\sum_{n=1}^{\infty} \mathfrak{t}_{(A_n,r)}$ have the same values on arguments $\mathfrak{g}_{\bar{c}}$ from the basis of the OOM. Thus, let $\bar{c} \in C_0$, $\bar{a} \in C^*$, and conclude

$$
\begin{aligned}
&\mathfrak{t}_{(\bigcup_{n=1}^{\infty} A_n, r)}\; \mathfrak{g}_{\bar{c}}(\bar{a}) = \\
&= P((\bigcup_{n=1}^{\infty} A_n, r)\,|\,\bar{c})\; \mathfrak{g}_{\bar{c}(\bigcup_{n=1}^{\infty} A_n, r)}\,(\bar{a}) \;=\; P((\bigcup_{n=1}^{\infty} A_n, r)\,|\,\bar{c})\; P(\bar{a}\,|\,\bar{c}\,(\bigcup_{n=1}^{\infty} A_n, r)) \\
&= P((\bigcup_{n=1}^{\infty} A_n, r)\bar{a}\,|\,\bar{c}) \;=\; \sum_{n=1}^{\infty} P((A_n, r)\bar{a}\,|\,\bar{c}) \\
&= \sum_{n=1}^{\infty} P((A_n, r)\,|\,\bar{c})\; P(\bar{a}\,|\,\bar{c}(A_n, r)) \\
&= \sum_{n=1}^{\infty} P((A_n, r)\bar{a}\,|\,\bar{c})\; \mathfrak{g}_{\bar{c}(A_n,r)}(\bar{a}) \;=\; \sum_{n=1}^{\infty} \mathfrak{t}_{(A_n,r)}\; \mathfrak{g}_{\bar{c}}(\bar{a}). \quad \square
\end{aligned}
$$

2. Like in the proof of *1.*, let $\bar{c} \in C_0$, $\bar{a} \in C^*$, and conclude

$$
\begin{aligned}
&\mathfrak{t}_{(A,r_1+r_2)}\; \mathfrak{g}_{\bar{c}}(\bar{a}) = \\
&= P((A, r_1 + r_2)\,|\,\bar{c})\; \mathfrak{g}_{\bar{c}(A,r_1+r_2)}\,(\bar{a}) \;=\; P((A, r_1 + r_2)\,|\,\bar{c})\; P(\bar{a}\,|\,\bar{c}(A, r_1 + r_2)) \\
&= P((A, r_1)(E, r_2)\,|\,\bar{c})\; P(\bar{a}\,|\,\bar{c}(A, r_1)(E, r_2)) \\
&= P((E, r_2)\,|\,\bar{c}(A, r_1))\; P((A, r_1)\,|\,\bar{c})\; P(\bar{a}\,|\,\bar{c}(A, r_1)(E, r_2)) \\
&= P((E, r_2)\,|\,\bar{c}(A, r_1))\; P((A, r_1)\,|\,\bar{c})\; \mathfrak{g}_{\bar{c}(A,r_1)(E,r_2)}(\bar{a}) \\
&= P((A, r_1)\,|\,\bar{c})\; \mathfrak{t}_{(E,r_2)}\; \mathfrak{g}_{\bar{c}(A,r_1)}\,(\bar{a}) \\
&= \mathfrak{t}_{(E,r_2)}\; P((A, r_1)\,|\,\bar{c})\; \mathfrak{g}_{\bar{c}(A,r_1)}(\bar{a}) \;=\; \mathfrak{t}_{(E,r_2)}\; \mathfrak{t}_{(A,r_1)}\; \mathfrak{g}_{\bar{c}}\,(\bar{a}). \quad \square
\end{aligned}
$$

Remark 3 *The denotation of a basic cylinder by a word is not unique. A basic cylinder* $\{X_{t_0} \in A_0, \ldots, X_{t_{n-1}} \in A_{n-1}\}$ *can be denoted by any word* $(E, t_0)(A_0, t_1 - t_0) \ldots (A_{n-2}, t_{n-1} - t_{n-2})(A_{n-1}, r)$*, where* $r > 0$*. This is not harmful in the sense that for two words* $\bar{a}(A, r), \bar{a}(A, r')$ *which differ only in their last time specification, it holds that* $\sigma\mathfrak{t}_{\bar{a}(A,r)}\mathfrak{g}_\varepsilon = \sigma\mathfrak{t}_{\bar{a}(A,r')}\mathfrak{g}_\varepsilon$ *(use Propositions 6 and 7 (2) for the simple proof).*

We conclude this section with an analog of Proposition 4, by providing necessary and sufficient conditions for a structure $(V, (e_j)_{j\in J}, (\tau_a)_{a\in C}, v_0)$ to specify a projective family of probability measures. Recall that the finite-dimensional marginal distributions of every stochastic process are such a projective family.

For the statement of the proposition we need some concepts and terminology. Let $(E, \mathfrak{B})$ be a measurable space. By $\mathcal{P}_0(\mathbb{R})$ we denote the set of finite subsets of $\mathbb{R}$. For $T = \{t_0, \ldots, t_{n-1}\} \in \mathcal{P}_0(\mathbb{R})$, let $\mathfrak{B}^T = \bigotimes_{t\in T} \mathfrak{B}_t$ be the product σ-algebra where every factor $\mathfrak{B}_t$ is equal to $\mathfrak{B}$. Furthermore, we shall use the symbol $\dot{\bigcup}$ to denote unions of *disjoint* sets, and use σ for the summation of basis vector coefficients like in (5), i.e. $\sigma \sum_i \alpha_i e_i = \sum_i \alpha_i$.

Proposition 8 *Let $(E, \mathfrak{B})$ be a measurable space. Let V be a vector space with a basis $(e_j)_{j\in J}$, let $(\tau_a)_{a\in C}$ be a family of linear operators on V, and let $v_0 \in V$. For $T = \{t_0, \ldots, t_{n-1}\} \in \mathcal{P}_0(\mathbb{R})$, where $t_0 < \cdots < t_{n-1}$, let the numerical function $P_T : \mathfrak{B}_{t_0} \otimes \cdots \otimes \mathfrak{B}_{t_{n-1}} \to \mathbb{R}$ be defined by*

$$P_T(A_{t_0} \times \cdots \times A_{t_{n-1}}) = \tag{8}$$
$$= \begin{cases} \sigma\tau_{(A_{n-1},1)}\tau_{(A_{n-2},t_{n-1}-t_{n-2})}\cdots\tau_{(A_1,t_2-t_1)}\tau_{(A_0,t_1-t_0)}\tau_{(E,t_0)}v_0, & \text{if } t_0 > 0, \\ \sigma\tau_{(A_{n-1},1)}\tau_{(A_{n-2},t_{n-1}-t_{n-2})}\cdots\tau_{(A_1,t_2-t_1)}\tau_{(A_0,t_1)} \qquad v_0, & \text{if } t_0 = 0. \end{cases}$$

Then $(P_T)_{T\in\mathcal{P}_0(\mathbb{R})}$ can be extended to a projective family $(E^T, \mathfrak{B}^T, P_T)_{T\in\mathcal{P}_0(\mathbb{R})}$ of probability measures on the cylinder sets, if and only if the following conditions are satisfied:

1. $\sigma v_0 = 1$,
2. $\sigma\tau_{(E,r)}e_j = \sigma e_j$ *for all* $r > 0$,
3. $\sigma\tau_{(A_{n-1},r_{n-1})}\cdots\tau_{(A_0,r_0)}v_0 \geq 0$ *for all* $(A_0, r_0)\ldots(A_{n-1}, r_{n-1}) \in C^*$,
4. $\tau_{(\dot{\bigcup}_{n=1}^{\infty} A_n, r)} = \sum_{n=1}^{\infty} \tau_{(A_n,r)}$ *for all* $r > 0$,
5. $\tau_{(A,r_1+r_2)} = \tau_{(A,r_2)} \circ \tau_{(E,r_1)}$ *for all* $r_1, r_2 > 0, A \in \mathfrak{B}$.

Note that conditions *1* – *3* correspond to the conditions known from the discrete case (Proposition 4), while *4* and *5* are the properties derived in Proposition 7.

Proof. The $\Longrightarrow$ direction is an easy exercise (re-use proofs of Propositions 6 and 7). We treat only the $\Longleftarrow$ case.

Step 1. We show that for $T = \{t_0, \ldots, t_{n-1}\}$ the numerical function P_T can be extended to a probability measure on $\mathfrak{B}^T$. We restrict our treatment to the case $t_0 = 0$ (the case $t_0 > 0$ can then be obtained in a straightforward way by considering $T' = \{0, t_0, \dots, t_{n-1}\}$). We follow the general scheme in [1] (§5) for constructing measures.

Step 1.1. We show that P_T can be extended to a pre-measure on the ring $\mathfrak{R}$ generated in $\mathfrak{B}^T$ by the sets $A_{t_0} \times \cdots \times A_{t_{n-1}} \in \mathfrak{B}_{t_0} \otimes \cdots \otimes \mathfrak{B}_{t_{n-1}}$ (a subset system is called a ring if it contains the empty set and is closed w.r.t. set complements and finite unions; a pre-measure on a ring is a non-negative numerical function which maps the empty set on 0 and is σ-additive).

Step 1.1.1. It is an elementary exercise to show every $A \in \mathfrak{R}$ can be represented by a finite disjoint union of sets from $\mathfrak{B}_{t_0} \otimes \cdots \otimes \mathfrak{B}_{t_{n-1}}$, i.e. $A = \dot{\bigcup}_{i=1,\dots,m} A_0^i \times \cdots \times A_{n-1}^i$ for suitable $A_j^i \in \mathfrak{B}$ (cf. [1] exercise §21 Nr. 1). We therefore can extend P_T on $\mathfrak{R}$ by putting

$$P_T(A) = \sum_{i=1,\dots,m} P_T(A_0^i \times \cdots \times A_{n-1}^i) \tag{9}$$

for a partition $A = \dot{\bigcup}_{i=1,\dots,m} A_0^i \times \cdots \times A_{n-1}^i$. We have to show that (9) is independent from the partition. Let $A = \dot{\bigcup}_{i=1,\dots,m} A_0^i \times \cdots \times A_{n-1}^i = \dot{\bigcup}_{j=1,\dots,k} B_0^j \times \cdots \times B_{n-1}^j$ be two partitions of A. Then there exists a common refinement

$$A \;=\; \dot{\bigcup}_{h=1,\dots,l} C_0^h \times \cdots \times C_{n-1}^h, \tag{10}$$

where every $C_0^h \times \cdots \times C_{n-1}^h$ is non-empty, and where $C_\nu^h \cap C_\nu^{h'} = \emptyset$ or $C_\nu^h = C_\nu^{h'}$ for all $h, h' \leq l; \nu \leq n-1$. Then, for $i \leq m$,

$$\begin{aligned} & A_0^i \times \cdots \times A_{n-1}^i = \\ & \quad = \dot{\bigcup}_{g_i=1,\dots,p_i} C_0^{g_i} \times \cdots \times C_{n-1}^{g_i} && (11) \\ & \quad =: \dot{\bigcup}_{j_0=1,\dots,m_0} C_0^{i,j_0} \times \cdots \times \dot{\bigcup}_{j_{n-1}=1,\dots,m_{n-1}} C_{n-1}^{i,j_{n-1}}, && (12) \end{aligned}$$

where $p_i \leq l$, (11) is a sub-collection of (10), and each C_ν^{i,j_ν} is one of the $C_\nu^{g_i}$. Exploiting condition *4* from the Proposition, we can now conclude

$$\begin{aligned} & \sum_{i=1,\dots,m} P_T(A_0^i \times \cdots \times A_{n-1}^i) = \\ & \quad = \sum_{i=1,\dots,m} P_T(\dot{\bigcup}_{j_0=1,\dots,m_0} C_0^{i,j_0} \times \cdots \times \dot{\bigcup}_{j_{n-1}=1,\dots,m_{n-1}} C_{n-1}^{i,j_{n-1}}) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1,\ldots,m} \sigma(\tau_{(\dot\bigcup_{j_{n-1}=1,\ldots,m_{n-1}} C^{i,j_{n-1}}_{n-1},1)} \circ \cdots \circ \tau_{(\dot\bigcup_{j_0=1,\ldots,m_0} C^{i,j_0}_0,t_1)}\ v_0) \\
&= \sum_{i=1,\ldots,m} \sigma \sum_{j_{n-1}=1,\ldots,m_{n-1}} \tau_{(C^{i,j_{n-1}}_{n-1},1)} \circ \cdots \circ \sum_{j_0=1,\ldots,m_0} \tau_{(C^{i,j_0}_0,t_1)}\ v_0 \\
&= \sigma \sum_{\substack{i=1,\ldots,m\\ j_0=1,\ldots,m_0\\ \cdots\\ j_{n-1}=1,\ldots,m_{n-1}}} \tau_{(C^{i,j_{n-1}}_{n-1},1)} \circ \cdots \circ \tau_{(C^{i,j_0}_0,t_1)}\ v_0 \\
&= \sigma \sum_{h=1,\ldots,l} \tau_{(C^h_{n-1},1)} \circ \cdots \circ \tau_{(C^h_0,t_1)}\ v_0 \qquad \text{[admissible since all occuring } C^{i,j_{n-1}}_{n-1} \times \cdots \times C^{i,j_0}_0 \text{ are disjoint.]}
\end{aligned}
$$

In a similar fashion we can show that $\sum_{j=1,\ldots,k} P_T(B^j_0 \times \cdots \times B^j_{n-1}) = \sigma \sum_{h=1,\ldots,l} \tau_{(C^h_{n-1},1)} \circ \cdots \circ \tau_{(C^h_0,0)}\ v_0$, which shows that (9) is independent from the partition, and concludes step 1.1.1.

Step 1.1.2. We show that P_T is σ-additive on $\mathfrak{R}$, i.e. for $A = \dot\bigcup_{i=1}^\infty A_i \in \mathfrak{R}$ it holds that $P_T(A) = \sum_{i=1}^\infty P_T(A_i)$. Let $A = \dot\bigcup_{j=1,\ldots,m} C^j_0 \times \cdots \times C^j_{n-1}$ be a finite partition of A, where all $C^j_\nu \in \mathfrak{B}$. Furthermore, we partition every A_i into $A_i = \dot\bigcup_{p=1,\ldots,s_i} D^{i,p}_0 \times \cdots \times D^{i,p}_{n-1} =: \dot\bigcup_{p=1,\ldots,s_i} D^{i,p}$ (where $s_i < \infty, D^{i,p} \in \mathfrak{B} \otimes \cdots \otimes \mathfrak{B}$). Let $C^j_\nu = \dot\bigcup_{k=1,\ldots,r_{\nu,j}} B^{k,j}_\nu$ (where $0 \le \nu \le n-1$, $r_{\nu,j} \le \infty, B^{k,j}_\nu \in \mathfrak{B}$) be a finite or infinite partition of C^j_ν, such that

(a) each $D^{i,p}_0 \times \cdots \times D^{i,p}_{n-1}$ is a disjoint union of sets of the form $B^{k_0,j}_0 \times \cdots \times B^{k_{n-1},j}_{n-1}$, namely, $D^{i,p}_0 \times \cdots \times D^{i,p}_{n-1} =: \dot\bigcup_{q=1}^{u_{i,p}} B^{i,p,q}_0 \times \cdots \times B^{i,p,q}_{n-1}$, where $u_{i,p} \le \infty$ and each $B^{i,p,q}_\nu$ is a set $B^{k_\nu,j}_\nu$ (where $1 \le j \le m$ and $1 \le k_\nu \le r_{\nu,j}$), and

(b) for all ν, k, k', j, j' it holds that $B^{k,j}_\nu \cap B^{k',j'}_\nu = \emptyset$ or $B^{k,j}_\nu = B^{k',j'}_\nu$.

It is not difficult to see that such a partition exists. (A way to obtain one is e.g. to first construct the coarsest common refinement $A = \dot\bigcup_x F^x_0 \times \cdots \times F^x_{n-1}$ of all of the sets $C^j_0 \times \cdots \times C^j_{n-1}$ $(j = 1, \ldots, m)$ and $D^{i,p}_0 \times \cdots \times D^{i,p}_{n-1}$ $(i \le \infty, p = 1, \ldots, s_i)$. Then, for every $0 \le \nu \le n-1$, further refine the sets F^x_ν to make them disjoint. Renaming yields the desired $B^{k,j}_\nu$.)

We now conclude

$$
\begin{aligned}
&P_T(\dot\bigcup_{i=1}^\infty A_i) = \\
&= P_T(\dot\bigcup_{j=1,\ldots,m} C^j_0 \times \cdots \times C^j_{n-1}) = \sum_{j=1,\ldots,m} \sigma \tau_{(C^j_{n-1},1)} \cdots \tau_{(C^j_0,t_1)}\ v_0
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j=1,\dots,m} \sigma\tau_{(\dot\bigcup_{k=1,\dots,r_{n-1,j}} B^{k,j}_{n-1},1)} \cdots \tau_{(\dot\bigcup_{k=1,\dots,r_{0,j}} B^{k,j}_{0},t_1)}\ v_0 \\
&= \sum_{j=1,\dots,m} \sigma(\sum_{k_{n-1}=1}^{r_{n-1,j}} \tau_{(B^{k_{n-1},j}_{n-1},1)}) \cdots (\sum_{k_0=1}^{r_{0,j}} \tau_{(B^{k_0,j}_{0},t_1)})\ v_0 \quad \text{[use condition } \mathit{4}\text{]} \\
&= \sum_{j=1,\dots,m} \sum_{\substack{k_0=1,\dots,r_{0,j} \\ \cdots \\ k_{n-1}=1,\dots,r_{n-1,j}}} \sigma\tau_{(B^{k_{n-1},j}_{n-1},1)} \cdots \tau_{(B^{k_0,j}_{0},t_1)}\ v_0 \quad \text{[exploit (b)]} \\
&= \sum_{i=1}^{\infty}\sum_{p=1}^{s_i}\sum_{q=1}^{u_{i,p}} \sigma\tau_{(B^{i,p,q}_{n-1},1)} \cdots \tau_{(B^{i,p,q}_{0},t_1)}\ v_0 \quad \text{[re-order according to (a); admissible because of cond. } \mathit{3}\text{]} \\
&= \sum_{i=1}^{\infty}\sum_{p=1}^{s_i} \sum_{\substack{q_0=1,\dots,u_{i,p,0} \\ \cdots \\ q_{n-1}=1,\dots,u_{i,p,n-1}}} \sigma\tau_{(B^{i,p,q_{n-1}}_{n-1},1)} \cdots \tau_{(B^{i,p,q_0}_{0},t_1)}\ v_0 \quad \text{[exploit again (b);} \\
&\quad B^{i,p,1}_{\nu},\dots,B^{i,p,u_{i,p,\nu}}_{\nu} \text{ is a repetition-free enumeration of } B^{i,p,1}_{\nu},\dots,B^{i,p,u_{i,p}}_{\nu}] \\
&= \sum_{i=1}^{\infty}\sum_{p=1}^{s_i} \sigma(\sum_{q_{n-1}=1}^{u_{i,p,n-1}} \tau_{(B^{i,p,q_{n-1}}_{n-1},1)}) \cdots (\sum_{q_0=1}^{u_{i,p,0}} \tau_{(B^{i,p,q_0}_{0},t_1)})\ v_0 \\
&= \sum_{i=1}^{\infty}\sum_{p=1}^{s_i} \sigma\tau_{(\dot\bigcup_{q_{n-1}=1}^{u_{i,p,n-1}} B^{i,p,q_{n-1}}_{n-1},1)} \cdots \tau_{(\dot\bigcup_{q_0=1}^{u_{i,p,0}} B^{i,p,q_0}_{n-1},t_1)}\ v_0 \quad \text{[apply again cond. } \mathit{4}\text{]} \\
&= \sum_{i=1}^{\infty}\sum_{p=1}^{s_i} \sigma\tau_{(D^{i,p}_{n-1},1)} \cdots \tau_{(D^{i,p}_{0},t_1)}\ v_0 \quad \text{[exploit that } D^{i,p}_0 = \dot\bigcup_{q_0=1}^{u_i,p,0} B^{i,p,q_0}_0] \\
&= \sum_{i=1}^{\infty}\sum_{p=1}^{s_i} P_T(D^{i,p}) \ = \ \sum_{i=1}^{\infty} P_T(A_i).
\end{aligned}
$$

This concludes step 1.1.

Step 1.2. We show that P_T is finite on $\mathfrak{R}$. Let $A = \dot\bigcup_{i=1}^{m} A^i_0 \times \cdots \times A^i_{n-1} \in \mathfrak{R}$, and conclude

$$
\begin{aligned}
&P_T(\dot\bigcup_{i=1}^{m} A^i_0 \times \cdots \times A^i_{n-1}) = \\
&\quad = \sum_{i=1}^{m} \sigma\tau_{(A^i_{n-1},1)} \cdots \tau_{(A^i_0,t_1)}\ v_0 \\
&\quad \le \sum_{i=1}^{m} \sigma\tau_{(E,1)} \cdots \tau_{(E,t_0)}\ v_0 \quad \text{[follows from conds. } \mathit{3} \text{ and } \mathit{4}\text{]} \\
&\quad = m. \quad \text{[follows from conds. } \mathit{1} \text{ and } \mathit{2}\text{]}
\end{aligned}
$$

This concludes step 1, since every finite pre-measure on a ring can be uniquely extended to a measure on the σ-algebra generated by the ring, and because $\mathfrak{R}$ generates $\mathfrak{B}^T$.

Step 2. We show that the family $(E^T, \mathfrak{B}^T, P_T)_{T \in \mathcal{P}_0(\mathbb{R})}$ is projective. We use the fact that the family $(P_T)_{T \in \mathcal{P}_0(\mathbb{R})}$ is projective if and only if

$$\begin{aligned} & P_{\{t_0,\ldots,t_\kappa,t_{\kappa+1},t_{\kappa+2},\ldots,t_{n-1}\}}(A_{t_0} \times \cdots \times A_{t_\kappa} \times E \times A_{t_{\kappa+2}} \times \cdots \times A_{t_{n-1}}) = \\ & = \; P_{\{t_0,\ldots,t_\kappa,t_{\kappa+2},\ldots,t_{n-1}\}}(A_{t_0} \times \cdots \times A_{t_\kappa} \times A_{t_{\kappa+2}} \times \cdots \times A_{t_{n-1}}) \end{aligned}$$

for all $n \geq 1, 0 \leq \kappa \leq n$. But this is a straightforward consequence of condition *5*.

This concludes the proof of Proposition 6. □

Under very general conditions, the projective family $(E^T, \mathfrak{B}^T, P_T)_{T \in \mathcal{P}_0(\mathbb{R})}$ can be extended to a probability measure on $\bigotimes_{t \in \mathbb{R}_{\geq 0}} \mathfrak{B}_t$, such that a stochastic process $(\Omega, \mathfrak{A}, P, (X_t)_{t \in \mathbb{R}_{\geq 0}})$ exists of which $(P_T)_{T \in \mathcal{P}_0(\mathbb{R})}$ yields the finite-dimensional marginal distributions (Kolmogorov extension theorem, cf. [11]).

A convex cone based version of condition *3* in Proposition 8, which would be analog to Prop. 5, remains to be worked out.

Like in the discrete case, we call OOMs $(\mathfrak{G}, (\mathfrak{e}_j)_{j \in J}, (\mathfrak{t}_A)_{A \in \mathfrak{B}}, \mathfrak{g}_\varepsilon)$ constructed via prediction functions, *canonical* OOMs. Note that in a general OOM $(V, (e_j)_{j \in J}, (\tau_a)_{a \in C}, v_0)$, V can have a vector space dimension greater than the dimension of the distribution it models. We call two OOMs *equivalent* if they model the same distribution. We call an OOM $\mathcal{A}$ *minimal-dimensional* if the vector space dimension of $\mathcal{A}$ is less or equal to the vector space dimension of $\mathcal{B}$ for any $\mathcal{B}$ that is equivalent to $\mathcal{A}$. It is easy to see that

1. canonical OOMs are minimal-dimensional,
2. equivalent, minimal-dimensional OOMs are isomorphic,
3. in a minimal-dimensional OOM $(V, (e_j)_{j \in J}, (\tau_a)_{a \in C}, v_0)$, V is spanned by the vectors $\{\tau_{\bar{a}} v_0 \,|\, \bar{a} \in C^*\}$.

Finally, we remark that discrete-time, arbitrary-valued, canonical OOMs $(\mathfrak{G}, (\mathfrak{e}_j)_{j \in J}, (\mathfrak{t}_A)_{A \in \mathfrak{B}}, \mathfrak{g}_\varepsilon)$ can be constructed from distributions in a similar fashion as continuous-time canonical OOMs. Such models describes finite-dimensional distributions of the kind $(X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1})$. The rigorous definition and construction are straightforward simplifications of the continuous-time case and left to the reader.

4 Stationary processes.

In this section we characterize stationary distributions in terms of their OOMs. Recall that a process is called stationary if the probabilities of elementary cylinders are invariant under temporal shift, which in our word notation can be expressed by the condition

$$\forall \bar{a} \in C^* \forall r > 0: \; P(\bar{a}) = P((E, r)\bar{a}). \tag{13}$$

Proposition 9 *Let $(V, (e_j)_{j \in J}, (\tau_a)_{a \in C}, v_0)$ be a minimal-dimensional OOM of a process. Then the process is stationary if and only if v_0 is invariant under all $\tau_{(E,r)}$, where $r > 0$. Without requiring minimal dimensionality, only the direction $\Longleftarrow$ holds.*

Proof. The $\Longleftarrow$ direction is obvious. For the $\Longrightarrow$ direction, assume that the process is stationary, i.e. that (13) holds. Since $(V, (e_j)_{j \in J}, (\tau_a)_{a \in C}, v_0)$ is minimal-dimensional, it is isomorphic to the canonical OOM $(\mathfrak{G}, (\mathfrak{e}_j)_{j \in J}, (\mathfrak{t}_{(a)})_{a \in C}, \mathfrak{g}_\varepsilon)$. It suffices therefore to show that $\mathfrak{t}_{(E,r)}\mathfrak{g}_\varepsilon = \mathfrak{g}_\varepsilon$ for all $r > 0$. Let $\bar{a} \in C^*$. Then, $\mathfrak{t}_{(E,r)}\mathfrak{g}_\varepsilon(\bar{a}) = P((E, r) \mid \bar{a})\mathfrak{g}_{(E,r)}(\bar{a}) = 1P(\bar{a} \mid (E, r)) = P(\bar{a}) = \mathfrak{g}_\varepsilon(\bar{a})$. Thus, $\mathfrak{t}_{(E,r)}\mathfrak{g}_\varepsilon = \mathfrak{g}_\varepsilon$. □

5 Decomposition of observable operators.

The family $(\tau_a)_{a \in C}$ of a continuous-time OOM is actually a doubly indexed family, namely, $(\tau_{(A,r)})_{A \in \mathfrak{B}, r \in \mathbb{R}_{>0}}$, which is rather unhandy. The topic of this section is how and when this doubly indexed family can be decomposed into two single-indexed families, which are simpler to handle and more revealing of the modeled distribution.

Sometimes such a decomposition is easy to obtain. Consider an OOM $(V, (e_j)_{j \in J}, (\tau_{(A,r)})_{A \in \mathfrak{B}, r \in \mathbb{R}_{>0}}, v_0)$, where every operator $\tau_{(E,r)}$ $(r > 0)$ is invertible. For every $A \in \mathfrak{B}$, define an operator η_A by putting

$$\eta_A = \tau^{-1}_{(E,r)} \circ \tau_{(A,r)}. \tag{14}$$

This definition does not depend on r. This can be seen as follows. Assume that $r > s > 0$. Then conclude

$$\begin{array}{lrcll}
 & \tau_{(A,r)} & = & \tau_{(A,r)} & \\
\Rightarrow & \tau_{(A,r)} & = & \tau_{(E,r-s)}\tau_{(E,s)}\tau^{-1}_{(E,s)}\tau_{(A,s)} & \text{[apply Prop. 7 (2)]} \\
\Rightarrow & \tau_{(E,r)}\tau^{-1}_{(E,r)}\tau_{(A,r)} & = & \tau_{(E,r)}\tau^{-1}_{(E,s)}\tau_{(A,s)} & \\
\Rightarrow & \tau^{-1}_{(E,r)}\tau_{(A,r)} & = & \tau^{-1}_{(E,s)}\tau_{(A,s)} &
\end{array}$$

Rename the operators $\tau_{(E,r)}$ into μ_r. Then for every operator $\tau_{(A,r)}$ it holds that

$$\tau_{(A,r)} = \mu_r \circ \eta_A, \tag{15}$$

because $\mu_r \eta_A = \tau_{(E,r)} \tau_{(E,r)}^{-1} \tau_{(A,r)} = \tau_{(A,r)}$. Thus, we have decomposed the family $(\tau_{(A,r)})_{A \in \mathfrak{B}, r \in \mathbb{R}_{>0}}$ into two single-indexed families $(\eta_A)_{A \in \mathfrak{B}}$ and $(\mu_r)_{r>0}$.

We remark in passing that if $\tau_{(E,r_0)}$ is invertible for some $r_0 > 0$, then the $\tau_{(E,s)}$ are invertible for all $s > 0$. This can be seen as follows. If $\tau_{(E,r_0)}$ is invertible, then all $\tau_{(E,nr_0)}$ are invertible, because $\tau_{(E,nr_0)} = \tau_{(E,r_0)} \circ \cdots \circ \tau_{(E,r_0)}$ (n times) as a consequence of Prop. 7 (*2*). Consider any $s > 0$. Choose n such that $nr_0 > s$. Then, again by Prop. 7 (*2*), $\tau_{(E,nr_0)} = \tau_{(E,nr_0 - s)} \tau_{(E,s)}$. Since the lhs. is invertible, both factors on the rhs. must be invertible, too. Thus, every $\tau_{(E,s)}$ is invertible.

Remark 4 *The decomposition (15) deserves more than a little comment. In the discrete case (section 2), we have intuitively interpreted OOMs as models of an observer's knowledge about an evolving system, and an observable operator $\mathfrak{t}_a$ as effecting evolution in the knowledge due to incoming observation a. In the continuous case, an observable operator $\mathfrak{t}_{(A,r)}$ describes changes of knowledge due to an observation of A, followed by a time interval r where no observation is made. When observable operators can be decomposed into operators $(\eta_A)_{A \in \mathfrak{B}}$ and $(\mu_r)_{r>0}$, the former describe "immediate jumps" in knowledge about the system's state due to an observation of A, while the latter describe the evolution of knowledge as time goes by with no observation available. In decomposed OOMs, we will call operators (η_A) "observation operators", while we shall refer to operators (μ_r) as "evolution operators".*

We now give a formal definition which covers both discrete and continuous-time OOMs.

Definition 4 *A decomposed OOM is a structure $(V, (e_j)_{j \in J}, (\eta_A)_{A \in \mathfrak{B}}, (\mu_r)_{r \in T_{>0}}, v_0)$, where $T_{>0} = \mathbb{N}_{>0}$ or $T_{>0} = \mathbb{R}_{>0}$, such that a stochastic process $(\Omega, \mathfrak{A}, P, (X_t)_{t \in T})$ exists (where $T = \mathbb{N}$ or $T = \mathbb{R}$), whose finite-dimensional marginal distributions can be computed by*

$$\begin{aligned} & P(X_0 \in A_0, X_{t_1} \in A_1, \ldots, X_{t_{n-1}} \in A_{n-1}) = \\ & = \sigma \eta_{A_{n-1}} \mu_{t_{n-1} - t_{n-2}} \cdots \eta_{A_{t_2}} \mu_{t_2 - t_1} \eta_{A_{t_1}} \mu_{t_1} \eta_{A_{t_0}} v_0. \end{aligned} \tag{16}$$

(The case $t_0 > 0$ is captured by putting $A_0 = E$, like in the previous section.)

Obviously, considering the case $T = \mathbb{R}$, if $(V, (e_j)_{j\in J}, (\eta_A)_{A\in\mathfrak{B}}, (\mu_r)_{r\in\mathbb{R}_{>0}}, v_0)$ is a decomposed OOM, then $(V, (e_j)_{j\in J}, (\mu_r \circ \eta_A)_{A\in\mathfrak{B}, r\in\mathbb{R}_{>0}}, v_0)$ is an OOM. Likewise, in the case $T = \mathbb{N}$, we obtain from a decomposed OOM $(V, (e_j)_{j\in J}, (\eta_A)_{A\in\mathfrak{B}}, (\mu_r)_{r\in\mathbb{N}_{>0}}, v_0)$ an ordinary discrete-time OOM $(V, (e_j)_{j\in J}, (\mu_1 \circ \eta_a)_{a\in E}, v_0)$.

In the remainder of this section I present a number of mixed results concerning the construction of decomposed OOMs. The first concerns finite-dimensional, finite-valued, discrete-time distributions. An OOM $\mathcal{A}$ of such a distribution can conveniently be written in matrix notation:

$$\mathcal{A} = (\mathbb{R}^m, (\tau_a)_{a\in E}, v_0),$$

where $m \in \mathbb{N}_{>0}$, the basis vectors $(e_j)_{1\le j\le m}$ are tacitly taken as the unit vectors and can therefore be omitted, E is a finite set of atomic observations, the operators τ_a are given by $m \times m$ real-valued matrices, and the σ-operation is left inner product with $(1, \ldots, 1)$, i.e.,

$$\sigma(x_1, \ldots, x_m)^T = (1, \ldots, 1) \cdot (x_1, \ldots, x_m)^T = x_1 + \cdots + x_m.$$

In matrix representation, equation (1) becomes

$$P(X_0 = a_{i_0}, \ldots, X_n = a_{i_n}) = \sigma\tau_{a_{i_n}} \cdots \tau_{a_{i_0}} v_0, \tag{17}$$

which here should be read as a sequence of matrix multiplications applied to v_0, concluded by the inner product with $(1, \ldots, 1)$. As was already mentioned in the introduction, the theory of such finite-dimensional OOMs in matrix representation has been elaborated in some detail. We will now see how from a finite-dimensional OOM $\mathcal{A} = (\mathbb{R}^m, (\tau_a)_{a\in E}, v_0)$ in matrix representation, one can construct a decomposed, equivalent OOM.

Let $E = \{a_1, \ldots, a_n\}$ have cardinality n. Define an $nm \times nm$ matrix $\tilde{\mu}$ by tiling the $nm \times nm$ array with n copies of each τ_{a_i} in a row, and define an nm vector $\tilde{v}_0$ by concatenating n weighted copies of v_0:

$$\tilde{\mu} = \left(\begin{array}{c|c|c} \tau_{a_1} & \cdots & \tau_{a_1} \\ \hline \tau_{a_2} & \cdots & \tau_{a_2} \\ \hline & \cdots & \\ \hline \tau_{a_n} & \cdots & \tau_{a_n} \end{array}\right) \qquad \tilde{v}_0 = \left(\begin{array}{c} \tau_{a_1}\ v_0 \\ \hline \tau_{a_2}\ v_0 \\ \hline \cdots \\ \hline \tau_{a_n}\ v_0 \end{array}\right) \tag{18}$$

Define $\tilde{\eta}_{a_i}$ to be the $nm \times nm$ matrix with zeroes everywhere except for ones at the diagonal positions $((n-1)i+1, (n-1)i+1), \ldots, ((n-1)i+m, (n-1)i+m)$. Finally, define $\tilde{\eta}_A = \sum_{a\in A} \tilde{\eta}_a$. It is a simple exercise to demonstrate that $(\mathbb{R}^{nm}, (e_j)_{j=1,\ldots,nm}, (\tilde{\eta}_A)_{A\subseteq E}, (\tilde{\mu}^t)_{t\in\mathbb{N}_{>0}}, \tilde{v}_0)$ is a decomposed OOM which is equivalent to $\mathcal{A}$, if one notes that $\tilde{\eta}_{a_{i_{k-1}}}\tilde{\mu}\cdots\tilde{\eta}_{a_{i_1}}\tilde{\mu}\tilde{\eta}_{a_{i_0}}\tilde{v}_0$ is the nm vector which has $\tau_{a_{i_{k-1}}}\cdots\tau_{a_{i_0}}v_0$ in its i_{k-1}th subcell of size m, and is zero elsewhere.

We now turn to the continuous-time case, and show how one can construct decomposed OOMs directly from the distribution by methods that are similar to the construction of canonical OOMs. We start with a denotation of basic cylinders by words that is slightly different from the one given in def. 3, in that time periods $r = 0$ are now allowed:

Definition 5 *(A variant of definition 3)*
Let

$$\begin{aligned}\tilde{C} &:= \{(A, r) \,|\, A \in \mathfrak{B}, r \in \mathbb{R}_{\geq 0}\},\\ \tilde{C}^* &:= \{(A_0, r_0)\ldots(A_{n-1}, r_{n-1}) \,|\, n \geq 1, A_i \in \mathfrak{B}, r_i \in \mathbb{R}_{\geq 0}\} \cup \{\varepsilon\}.\end{aligned}$$

We use symbols $\tilde{a}, \tilde{b}, \tilde{c}$ for elements of $\tilde{C}$, etc. Use elements of $\tilde{C}^*$ to denote basic cylinders, as follows:

1. *ε denotes the basic cylinder $\{X_0 \in E\}$,*
2. *A word $\bar{\tilde{a}} \in \tilde{C}^*$ of the form*

$$\begin{aligned}\bar{\tilde{a}} &= \\ &= (A_0^1, 0)\ldots(A_0^{i_0-1}, 0)(A_0^{i_0}, r_0)\\ &\quad (A_1^1, 0)\ldots(A_1^{i_1-1}, 0)(A_1^{i_1}, r_1)\\ &\quad \ldots\\ &\quad (A_{n-1}^1, 0)\ldots(A_{n-1}^{i_{n-1}-1}, 0)(A_{n-1}^{i_{n-1}}, r_{n-1}),\end{aligned}$$

   where $n \geq 1$; $r_0, \ldots, r_{n-1} > 0$; $i_0, \ldots, i_{n-1} \geq 1$, denotes the basic cylinder

$$\begin{aligned}&\{X_0 \in A_0^1 \cap \cdots \cap A_0^{i_0},\ X_{r_0} \in A_1^1 \cap \cdots \cap A_1^{i_1},\ \ldots,\\ &X_{r_0+\cdots+r_{n-2}} \in A_{n-1}^1 \cap \cdots \cap A_{n-1}^{i_{n-1}}\}.\end{aligned}$$

That is, if a word contains "blocks" of consecutive zero-time $(A^j, 0)$, the intersection of all concerned events A^j is taken in the basic cylinder.

We write $P(\bar{\tilde{a}})$ for the probability of the basic cylinder denoted by $\bar{\tilde{a}}$. Probabilities of concatenations $P(\bar{\tilde{b}}\bar{\tilde{a}})$ and conditional probabilities $P(\bar{\tilde{b}} \mid \bar{\tilde{a}})$ are defined in the obvious way.

The original statement and proof of Proposition 6 can be repeated with the only alteration that time indices $r = 0$ are now allowed. For convenience, we re-state this slight variant of Proposition 6:

Proposition 10 *(A variant of Prop. 6). Let $(\Omega, \mathfrak{A}, P, (X_t)_{t \in \mathbb{R}_{\geq 0}})$ be a stochastic process with values in $(E, \mathfrak{B})$. Then there exist a real vector space $\tilde{\mathfrak{G}}$, a basis $(\tilde{\mathfrak{e}}_j)_{j \in J}$ of $\tilde{\mathfrak{G}}$, a vector $\tilde{\mathfrak{g}}_\varepsilon \in \tilde{\mathfrak{G}}$, and a family of linear operators $(\tilde{\mathfrak{t}}_{(A,r)})_{A \in \mathfrak{B}, r \in \mathbb{R}_{\geq 0}}$, such that the probability of the basic cylinder denoted by $\bar{\tilde{a}} = \tilde{a}_0 \cdots \tilde{a}_{n-1} = (A_0, r_0) \cdots (A_{n-1}, r_{n-1})$ can be computed in the following way:*

$$P(\bar{\tilde{a}}) = \sigma_{(\tilde{\mathfrak{e}}_j)_{j \in J}}(\tilde{\mathfrak{t}}_{(A_{n-1}, r_{n-1})} \cdots \tilde{\mathfrak{t}}_{(A_1, r_1)} \tilde{\mathfrak{t}}_{(A_0, r_0)} \tilde{\mathfrak{g}}_\varepsilon). \tag{19}$$

Since the prediction functions $\mathfrak{g}_{\bar{\tilde{c}}}$ used in the construction of $\tilde{\mathfrak{G}}$ are a superset of the prediction functions that span our accustomed vector space $\mathfrak{G}$, the vector space dimension of $\tilde{\mathfrak{G}}$ is greater or equal to that of $\mathfrak{G}$.

The desired construction of a decomposed OOM is now obtained as an obvious corollary.

Proposition 11 *Let $(\tilde{\mathfrak{G}}, (\tilde{\mathfrak{e}}_j)_{j \in J}, (\tilde{\mathfrak{t}}_{(A,r)})_{A \in \mathfrak{B}, r \in \mathbb{R}_{\geq 0}}, \tilde{\mathfrak{g}}_\varepsilon)$ be constructed from the distribution of $(\Omega, \mathfrak{A}, P, (X_t)_{t \in \mathbb{R}_{\geq 0}})$ according to Prop. 10. Then $(\tilde{\mathfrak{G}}, (\tilde{\mathfrak{e}}_j)_{j \in J}, (\tilde{\mathfrak{t}}_{(A,0)})_{A \in \mathfrak{B}}, (\tilde{\mathfrak{t}}_{(E,r)})_{r \in \mathbb{R}_{>0}}, \tilde{\mathfrak{g}}_\varepsilon)$ is a decomposed OOM which models the same distribution.*

Summing up, in this section we have seen that for every OOM $\mathcal{A}$ one can find an equivalent, decomposed OOM $\mathcal{B}$, albeit possibly of a higher dimension. If the mapping μ associated with $\mathcal{A}$ is invertible, then a decomposed OOM $\mathcal{B}$ can be provided without increasing dimension. It is easy to find discrete, finite-dimensional, minimal-dimensional examples where μ is not invertible, but where nevertheless a decomposition does not raise dimension. It is also easy to find minimal-dimensional examples where every decomposition raises the dimension: for instance, let μ be rank deficient, but some τ_a have full rank; then no η_a can satisfy the requirement $\mu \eta_a = \tau_a$. Outside such finite-dimensional rank considerations, nothing is known about conditions when decomposition implies an increase of dimensionality.

6 Conclusion.

Textbooks on stochastic processes are by and large organized along specific classes of processes: e.g., Markov processes, martingales, processes with independent increments, etc. OOM theory contributes to this organization of the field by offering novel, interesting classifications which arise from algebraic properties. For example, finite-dimensional processes are worth a separate treatment, as would be processes where the evolution operator μ is invertible.

Conversely, known classes of processes can be characterized by their algebraic properties. We have seen an instance of this in section 4, where stationary processes have been characterized. Historically, it was the (solved) task of algebraically characterizing finite-dimensional hidden Markov processes which gave rise to OOM theory. Besides that, it is not difficult to characterize processes with i.i.d. random variables, or Markov processes – it may amuse the reader to do so. Other known classes of processes remain to be investigated.

OOM theory is young, and there are many open problems. Some of them, which are connected to themes addressed in the present article, are the following:

- Decide (in the discrete-time, discrete-valued case) whether a candidate structure $(\mathbb{R}^m, (\tau_a)_{a \in E}, v_0)$ is a valid OOM, i.e., satisfies the conditions from proposition 4.
- Given a reduced OOM, when can it be decomposed without raising the dimension?
- In OOMs derived from finite-dimensional hidden Markov processes, the semigroup $(\mu_r)_{r>0}$ is a Markov semigroup (which gives rise to the underlying Markov process). Furthermore, the algebraic properties of $(\mu_r)_{r>0}$ are tightly coupled to the ergodic properties of the hidden Markov process. This motivates the question of how the ergodic properties of some process can be connected with the algebraic properties of the corresponding semigroup $(\mu_r)_{r>0}$ of evolution operators.

Acknowledgements. Many thanks go to Vladislav Tadić for careful proofreading and valuable suggestions for transparent formulations. The results described in this paper were obtained while the author worked under a postdoctoral grant from GMD.

References

[1] H. Bauer. *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. de Gruyter, Berlin/New York, 3 edition, 1978. English translation: Probability Theory and Elements of Measure Theory, New York: Holt, Rinehart & Winston, 1972.

[2] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences.* Academic Press, 1979.

[3] D. Blackwell and L. Koopmans. On the identifiability problem for functions of finite Markov chains. *Annals of Mathematical Statistics*, 38:1011–1015, 1957.

[4] S.W. Dharmadhikari. Functions of finite Markov chains. *Annals of Mathematical Statistics*, 34:1022–1032, 1963.

[5] S.W. Dharmadhikari. Sufficient conditions for a stationary process to be a function of a finite Markov chain. *Annals of Mathematical Statistics*, 34:1033–1041, 1963.

[6] S.W. Dharmadhikari. A characterization of a class of functions of finite Markov chains. *Annals of Mathematical Statistics*, 36:524–528, 1965.

[7] M. Fox and H. Rubin. Functions of processes with Markovian states. *Annals of Mathematical Statistics*, 39(3):938–946, 1968.

[8] M. Fox and H. Rubin. Functions of processes with Markovian states II. *Annals of Mathematical Statistics*, 40(3):865–869, 1969.

[9] M. Fox and H. Rubin. Functions of processes with Markovian states III. *Annals of Mathematical Statistics*, 41(2):472–479, 1970.

[10] E.J. Gilbert. On the identifiability problem for functions of finite Markov chains. *Annals of Mathematical Statistics*, 30:688–697, 1959.

[11] R.M. Gray. *Probability, Random Processes, and Ergodic Properties.* Springer Verlag, 1988.

[12] A. Heller. On stochastic processes derived from Markov chains. *Annals of Mathematical Statistics*, 36:1286–1291, 1965.

[13] M. Iosifescu and R. Theodorescu. *Random Processes and Learning*, volume 150 of *Die Grundlagen der mathematischen Wissenschaften in Einzeldarstellungen.* Springer Verlag, 1969.

[14] H. Ito. *An algebraic study of discrete stochastic systems.* Phd thesis, Dpt. of Math. Engineering and Information Physics, Faculty of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo, 1992. ftp'able from http://kuro.is.sci.toho-u.ac.jp:8080/english/D/.

[15] H. Ito, S.-I. Amari, and K. Kobayashi. Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE transactions on information theory*, 38(2):324–333, 1992.

[16] H. Jaeger. Discrete-time, discrete-valued observable operator models: a tutorial. GMD Report 42, GMD, Sankt Augustin, 1998. http://www.gmd.de/People/ Herbert.Jaeger/Publications.html.

[17] H. Jaeger. Observable operator models for discrete stochastic time series. *Neural Computation, to appear*, 1999. Download from http://www.gmd.de/People/Herbert.Jaeger/.

[18] L.A. Zadeh. The concept of system, aggregate, and state in system theory. In L.A. Zadeh and E. Polak, editors, *System Theory*, volume 8 of *Inter-University Electronics Series*, pages 3–42. McGraw-Hill, New York, 1969.