Rolf Gohm

Prediction Errors and

Completely Positive Maps

Abstract. We introduce the concept of an adapted isometry which is an operator-theoretic characterization of the time evolution of a stationary stochastic process adapted to a filtration. Using a product decomposition of an adapted isometry it is shown that prediction errors with respect to the filtration are related to a sequence of completely positive maps. Asymptotic properties of this correspondence are studied. In a special case the computations can be simplified by stochastic matrices.

Mathematics Subject Classification (1995): 28D05, 46L55, 46L57, 60G25

1. Introduction

In this paper we want to describe a surprising link between prediction errors for a stationary stochastic process and completely positive maps on corresponding spaces of operators. We consider discrete time steps. The process is assumed to be adapted to a filtration, and prediction means to compute the conditional expectation of a random variable of the process with respect to a sub-σ-algebra in the filtration corresponding to some earlier time. Details are given below.

Ernst-Moritz-Arndt-Universität, Institut für Mathematik und Informatik
Friedrich-Ludwig-Jahn-Str. 15 a, D-17489 Greifswald, Germany
email: gohm@mail.uni-greifswald.de

From stationarity there is an isometric time shift. In Section 2 we propose the concept of an "adapted isometry" which captures the essential features of such a time shift with respect to the filtration in a purely operator-theoretic way. Then a product decomposition of an adapted isometry is constructed which makes possible the step-by-step analysis to follow.

In Section 3 we define a sequence of completely positive maps acting on spaces of trace class operators in such a way that certain non-linear prediction errors of the process with respect to the filtration can be expressed by products of these maps. In other words, the evolution of the prediction errors in time can be described by a sort of dynamics which is well-known in quantum theory as an irreversible dynamics of mixed states. This correspondence is our main observation. As a first application we relate asymptotic behaviour of the process (determinism) with asymptotic properties of this dynamics (absorbing vector states).

Our theory applies in the same way to usual (commutative) stochastic processes and to non-commutative stochastic processes of operators. While an application to non-commutative processes seems very promising it requires some preparation and we therefore decided to postpone it to later work. Instead we analyse in Section 4 what we get for a commutative process with finitely many values. This restriction leads to a remarkable simplification: With respect to a suitable basis in the (now finite-dimensional) spaces of trace class operators the completely positive maps constructed in Section 3 turn out to be stochastic matrices. This allows us to write down combinatorial versions of our results. Using well-known facts about stochastic matrices the asymptotic behaviour of such processes can be quickly determined.

The paper relies mainly on explicit computations and is essentially self-contained. For concepts from operator theory we give [6] as a general reference.

2. Adapted Isometries

In this paper we consider Hilbert spaces which are all assumed to be complex and separable. Let $\{\mathcal{K}_n\}_{n=0}^{\infty}$ be a sequence of such Hilbert spaces and let $\{\Omega_n\}_{n=0}^{\infty}$ be a sequence of unit vectors so that $\Omega_n \in \mathcal{K}_n$ for all n. Then there is an infinite tensor product $\tilde{\mathcal{K}} = \bigotimes_{n=0}^{\infty} \mathcal{K}_n$ along the given sequence of unit vectors (cf. [6], 11.5.29). There is a distinguished unit vector $\tilde{\Omega} = \bigotimes_{n=0}^{\infty} \Omega_n \in \tilde{\mathcal{K}}$. Further we consider the subspaces $\mathcal{K}_{[m,n]} = \bigotimes_{j=m}^{n} \mathcal{K}_j \quad (m \leq n)$ of $\tilde{\mathcal{K}}$ where $\eta \in \mathcal{K}_{[m,n]}$ is identified with $\bigotimes_{j=0}^{m-1} \Omega_j \otimes \eta \otimes \bigotimes_{j=n+1}^{\infty} \Omega_j \in \tilde{\mathcal{K}}$. Then $\tilde{\mathcal{K}}$ is the closure of $\bigcup_{n=0}^{\infty} \mathcal{K}_{[0,n]}$. An operator $a \in \mathcal{B}(\mathcal{K}_{[m,n]})$ ($\mathcal{B}$ denotes bounded linear operators) is identified with $\mathbb{1}_{[0,m-1]} \otimes a \otimes \mathbb{1}_{[n+1,\infty)} \in \mathcal{B}(\tilde{\mathcal{K}})$.

Definition 2.1. *An isometry $\tilde{v} \in \mathcal{B}(\tilde{\mathcal{K}})$ is called adapted (with respect to $\{\mathcal{K}_n, \Omega_n\}_{n=0}^{\infty}$) if $\tilde{v}\, \mathcal{K}_{[0,n]} \subset \mathcal{K}_{[0,n+1]}$ for all $n \in \mathbb{N}_0$ and $\tilde{v}\tilde{\Omega} = \tilde{\Omega}$.*

This terminology is motivated by probability theory. Let us indicate how adapted isometries arise from stochastic processes:

Let (Ω, Σ, μ) be a probability space and $\{\Sigma_n\}_{n=0}^\infty$ a sequence of independent sub-σ-algebras. Denote by $\Sigma_{[m,n]}$ the σ-algebra generated by all Σ_j for $m \le j \le n$. In particular we get a filtration $\{\Sigma_{[0,n]}\}_{n=0}^\infty$ of increasing sub-σ-algebras. We assume that they together generate Σ.

There is a (discrete time) stationary stochastic process on (Ω, Σ, μ) determined by the sub-σ-algebra Σ_0 (representing time 0) together with a (not necessarily invertible) measurable and μ-preserving transformation $\tilde{\tau} : \Omega \to \Omega$. Namely, functions which are measurable with respect to $\tilde{\tau}^{-n}\Sigma_0$ may be interpreted as random variables of the process at time n. The process given by Σ_0 and $\tilde{\tau}$ is adapted with respect to the filtration $\{\Sigma_{[0,n]}\}_{n=0}^\infty$ if $\tilde{\tau}^{-n}\Sigma_0 \subset \Sigma_{[0,n]}$ for all n. This is the case if $\tilde{\tau}^{-1}\Sigma_{[0,n]} \subset \Sigma_{[0,n+1]}$ for all n, and it is no serious restriction of generality to consider only this setting.

Because $\tilde{\tau}$ is μ-preserving it induces an isometry $\tilde{v}$ on the square integrable random variables by $\tilde{v}\xi(\omega) := \xi(\tilde{\tau}\omega)$ for $\xi \in L^2(\Omega, \Sigma, \mu)$. Assume that $L^2(\Omega, \Sigma, \mu)$ is separable as a Hilbert space and that $\tilde{\tau}^{-1}\Sigma_{[0,n]} \subset \Sigma_{[0,n+1]}$ for all n. Then it is easy to check that $\tilde{v}$ is an adapted isometry with respect to $\{L^2(\Omega, \Sigma_n, \mu), 1_n\}_{n=0}^\infty$, where 1_n denotes the constant unit function considered as an element of $L^2(\Omega, \Sigma_n, \mu)$: First use the well-known fact that independence of σ-algebras implies a tensor product decomposition of the corresponding L^2-spaces and then translate the properties of $\tilde{\tau}$ into properties of $\tilde{v}$.

Let us remark here that any adapted isometry may be constructed by probabilistic means if we include non-commutative processes in the sense of

[1]. Here the probability space and its sub-σ-algebras are replaced by a unital *-algebra with a state and its unital *-subalgebras, the transformation $\tilde{\tau}$ by a unital *-homomorphism preserving the state and the space $L^2(\Omega, \Sigma, \mu)$ by the corresponding GNS-Hilbert space. If we then use a notion of independence based on tensor products then we may repeat the arguments above in this more general setting. In particular if we start with an adapted isometry $\tilde{v}$ as in Definition 2.1 we may use $\mathcal{B}(\tilde{\mathcal{K}})$ as an algebra, $\tilde{\Omega}$ as a vector state, subalgebras $\mathcal{B}(\mathcal{K}_{[m,n]})$ and the homomorphism defined by $x \mapsto \tilde{v}\, x\, \tilde{v}^*$. From this we can reconstruct $\tilde{v}$ as the adapted isometry corresponding to a non-commutative stationary stochastic process. Because we shall not explicitely consider non-commutative stochastic processes in this paper we do not give more details.

Let us now analyse the structure of an adapted isometry from an operator-theoretic point of view.

Proposition 2.2. *Let a sequence* $\{\mathcal{K}_n, \Omega_n\}_{n=0}^{\infty}$ *be given as above.*
If $\{u_n\}_{n=1}^{\infty} \subset \mathcal{B}(\tilde{\mathcal{K}})$ *is a sequence of unitaries so that*

$$\text{for all } n \geq 1 \quad u_n \in \mathcal{B}(\mathcal{K}_{[0,n]}),\ u_n \tilde{\Omega} = \tilde{\Omega},$$

$$\text{for all } n \geq 2 \quad u_n\,|_{\mathcal{K}_{[0,n-2]}} = \mathbb{1}\,|_{\mathcal{K}_{[0,n-2]}},$$

then we can define an adapted isometry by $\tilde{v} := stop - \lim_{n\to\infty} u_1 u_2 \ldots u_n$ *(here "stop" denotes the strong operator topology).*
Conversely if $\dim \mathcal{K}_n < \infty$ *for all* n *then any adapted isometry can be written in that way.*

Proof: If $\xi \in \mathcal{K}_{[0,m-1]}$ then it is fixed by any u_n with $n \geq m+1$. Therefore $\tilde{v}\xi = \lim_{n\to\infty} u_1 u_2 \ldots u_n \xi = u_1 u_2 \ldots u_m \xi \in \mathcal{K}_{[0,m]}$. By approximation the limit exists for all $\xi \in \tilde{\mathcal{K}}$.

To prove the converse let $\tilde{v}$ be an adapted isometry. For all $n \geq 1$ consider the isometry $\tilde{v}_n : \mathcal{K}_{[0,n-1]} \to \mathcal{K}_{[0,n]}$ given by the restriction of $\tilde{v}$ to $\mathcal{K}_{[0,n-1]}$. By assumption we deal here with finite dimensional spaces and therefore by dimension arguments there is an extension of $\tilde{v}_n$ to a unitary $\tilde{u}_n \in \mathcal{B}(\mathcal{K}_{[0,n]})$. Now define $u_n := \tilde{u}_{n-1}^{-1}\tilde{u}_n$ (with $\tilde{u}_0 := \mathbb{1}$). We have $u_n \in \mathcal{B}(\mathcal{K}_{[0,n]})$ and from $\tilde{v}\tilde{\Omega} = \tilde{\Omega}$ we also get $u_n\tilde{\Omega} = \tilde{\Omega}$. If $n \geq 2$ then because of $\tilde{u}_{n-1}\,|_{\mathcal{K}_{[0,n-2]}} = \tilde{v}\,|_{\mathcal{K}_{[0,n-2]}} = \tilde{u}_n\,|_{\mathcal{K}_{[0,n-2]}}$ we find that $u_n\,|_{\mathcal{K}_{[0,n-2]}} = \mathbb{1}\,|_{\mathcal{K}_{[0,n-2]}}$. □

We add some remarks. First: If we drop the condition of finite dimensionality for the converse direction then an inspection of the proof above shows that it may be necessary to enlarge the spaces $\mathcal{K}_n$ in order to proceed. This shows that any adapted isometry can at least be embedded into an adapted isometry of the product type above. Second: The condition $u_n\,|_{\mathcal{K}_{[0,n-2]}} = \mathbb{1}\,|_{\mathcal{K}_{[0,n-2]}}$ is implied by the more convenient condition $u_n \in \mathcal{B}(\mathcal{K}_{[n-1,n]})$. It is an interesting question how representability with this stronger condition restricts the class of associated stochastic processes. We shall not pursue this question here. Instead for the rest of this paper we shall assume that the adapted isometry is given by a sequence of unitaries $\{u_n\}_{n=1}^{\infty}$ satisfying even the stronger condition and ask how we can use this step-by-step information to analyse the associated stochastic process.

3. Main results

Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces. Denote by $\mathcal{T}(\ldots)$ the space of trace class operators with Tr the trace functional and $Tr_{\mathcal{H}} : \mathcal{T}(\mathcal{H} \otimes \mathcal{K}) \to \mathcal{T}(\mathcal{K})$ the partial trace obtained by evaluating the trace only on $\mathcal{H}$. To any isometry $v : \mathcal{H} \to \mathcal{H} \otimes \mathcal{K}$ we can associate the operator $D_v : \mathcal{T}(\mathcal{H}) \to \mathcal{T}(\mathcal{K})$ given by $\rho \mapsto Tr_{\mathcal{H}}(v\rho v^*)$. Then D_v is a completely positive and Tr-preserving map. In the physical literature such maps are important because they define time evolutions for mixed quantum states or density matrices (cf. [2]). Mathematically this is the set of positive trace class operators with unit trace which we denote by $\mathcal{T}_1^+(\ldots)$. Because Dirac's notation is useful for the computations to follow we decided to adopt it together with the physical convention of scalar products linear in the second component.

Given a unit vector $\Omega_{\mathcal{K}} \in \mathcal{K}$ we can consider $\mathcal{H}$ as a subspace of $\mathcal{H} \otimes \mathcal{K}$ by $\mathcal{H} \simeq \mathcal{H} \otimes \Omega_{\mathcal{K}} \subset \mathcal{H} \otimes \mathcal{K}$. Let p be the corresponding orthogonal projection. We may interpret v as an isometric dilation of the contraction $pv \in \mathcal{B}(\mathcal{H})$ (cf. [5], chapter VI). The following Lemma shows that D_v encodes information about the defect arising in this dilation procedure.

Lemma 3.1. *Assume $\xi, \xi' \in \mathcal{H}$. Then*

(a) $Tr_{\mathcal{H}}(|v\xi'><v\xi|) \; = \; D_v(|\xi'><\xi|)$,

(b) $<pv\xi, pv\xi'> = <\Omega_{\mathcal{K}}, D_v(|\xi'><\xi|)\, \Omega_{\mathcal{K}}>$.

Proof: (a) is immediate from the definition of D_v. To prove *(b)* we choose an orthonormal basis $\{\epsilon_i\}$ of $\mathcal{K}$ with $\epsilon_1 = \Omega_{\mathcal{K}}$. Then $v\xi = \sum_i \xi_i \otimes \epsilon_i$ and

$v\xi' = \sum_i \xi_i' \otimes \epsilon_i$ with $\{\xi_i\}$, $\{\xi_i'\} \subset \mathcal{H}$. We conclude that

$$< \Omega_{\mathcal{K}}, D_v(|\xi'><\xi|)\, \Omega_{\mathcal{K}} > = < \epsilon_1, Tr_{\mathcal{H}}\,(|v\xi'><v\xi|)\, \epsilon_1 >$$

$$= < \epsilon_1, \left(\sum_{i,j} <\xi_i, \xi_j'> |\epsilon_j><\epsilon_i| \right) \epsilon_1 > = <\xi_1, \xi_1'> = <pv\xi, pv\xi'> .$$

□

Lemma 3.2. *Let $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2$ be Hilbert spaces, $u : \mathcal{H}_0 \otimes \mathcal{H}_1 \to \mathcal{H}_0 \otimes \mathcal{H}_1$ unitary, $w : \mathcal{H}_1 \to \mathcal{H}_1 \otimes \mathcal{H}_2$ isometric and $v : \mathcal{H}_0 \otimes \mathcal{H}_1 \to \mathcal{H}_0 \otimes \mathcal{H}_1 \otimes \mathcal{H}_2$ defined as $v = (u \otimes \mathbb{1}_{\mathcal{H}_2})(\mathbb{1}_{\mathcal{H}_0} \otimes w)$. Then $D_v = D_w \circ Tr_{\mathcal{H}_0}$.*

Proof: For $\rho \in \mathcal{T}(\mathcal{H}_0 \otimes \mathcal{H}_1)$ we find that $D_w \circ Tr_{\mathcal{H}_0}(\rho)$

$$= Tr_{\mathcal{H}_1} \left(w\, Tr_{\mathcal{H}_0}(\rho)\, w^* \right) = Tr_{\mathcal{H}_0 \otimes \mathcal{H}_1} \left(\mathbb{1}_{\mathcal{H}_0} \otimes w \quad \rho \quad \mathbb{1}_{\mathcal{H}_0} \otimes w^* \right)$$

$$= Tr_{\mathcal{H}_0 \otimes \mathcal{H}_1} \left(u \otimes \mathbb{1}_{\mathcal{H}_2}\ \mathbb{1}_{\mathcal{H}_0} \otimes w \quad \rho \quad \mathbb{1}_{\mathcal{H}_0} \otimes w^*\ u^* \otimes \mathbb{1}_{\mathcal{H}_2} \right)$$

$$= Tr_{\mathcal{H}_0 \otimes \mathcal{H}_1}(v \rho v^*) = D_v(\rho).$$

□

We want to use these results for the analysis of adapted isometries. Let $\tilde{v} \in \mathcal{B}(\tilde{\mathcal{K}})$ be an adapted isometry (with respect to $\{\mathcal{K}_n, \Omega_n\}_{n=0}^{\infty}$, see Definition 2.1) and let $\{u_n\}_{n=1}^{\infty}$ be an associated sequence of unitaries as in Proposition 2.2. We assume that $u_n \in \mathcal{B}(\mathcal{K}_{[n-1,n]})$ for all n. Now define isometries $v_n : \mathcal{K}_{n-1} \to \mathcal{K}_{[n-1,n]} = \mathcal{K}_{n-1} \otimes \mathcal{K}_n$ as restrictions of u_n to $\mathcal{K}_{n-1}$. Using the procedure above we get a sequence of Tr-preserving completely positive maps $D_n := D_{v_n} : \mathcal{T}(\mathcal{K}_{n-1}) \to \mathcal{T}(\mathcal{K}_n)$, $n \geq 1$. Note that $D_n(|\, \Omega_{n-1} >< \Omega_{n-1}\, |) = |\, \Omega_n >< \Omega_n\, |$ (because $u_n \tilde{\Omega} = \tilde{\Omega}$). For the subspace $\mathcal{K}_{[m,n]}$ of $\tilde{\mathcal{K}}$ let us denote by $p_{[m,n]}$ resp. $Tr_{[m,n]}$ the corresponding orthogonal projection resp. partial trace.

Theorem 3.3. *Assume $\xi, \xi' \in \mathcal{K}_0$. Then for all $n \in \mathbb{N}_0$:*

(a) $Tr_{[0,n]}(|\tilde{v}^{n+1}\xi'><\tilde{v}^{n+1}\xi|) \;=\; D_{n+1}D_n \dots D_1(|\xi'><\xi|)$

(b) $<p_{[0,n]}\tilde{v}^{n+1}\xi,\, p_{[0,n]}\tilde{v}^{n+1}\xi'> = <\Omega_{n+1}, D_{n+1}D_n \dots D_1(|\xi'><\xi|)\Omega_{n+1}>$

Proof: *(b)* follows from *(a)* in the same way as shown in the proof of Lemma 3.1. To prove *(a)* we proceed by induction.

The case $n = 0$ is given by Lemma 3.1 (note that $\tilde{v}\,|_{\mathcal{K}_0} = v_1$). Now for some $n \geq 1$ assume that $Tr_{[0,n-1]}(|\,\tilde{v}^n\xi'><\tilde{v}^n\xi\,|) = D_n \dots D_1(|\,\xi'><\xi\,|)$. We have $\tilde{v}^{n+1}\xi = u_1 \dots u_n u_{n+1}\tilde{v}^n\xi = u_1 \dots u_n v_{n+1}\tilde{v}^n\xi$ (and the same for ξ'). Applying Lemma 3.2 with $\mathcal{H}_0 = \mathcal{K}_{[0,n-1]}$, $\mathcal{H}_1 = \mathcal{K}_n$, $\mathcal{H}_2 = \mathcal{K}_{n+1}$, $u = u_1 \dots u_n$, $w = v_{n+1}$, $v = \tilde{v}\,|_{\mathcal{H}_0 \otimes \mathcal{H}_1}$ we get

$$Tr_{[0,n]}(|\,\tilde{v}^{n+1}\xi'><\tilde{v}^{n+1}\xi\,|) \;=\; D_v(|\,\tilde{v}^n\xi'><\tilde{v}^n\xi\,|)$$

$$= D_{n+1}(Tr_{[0,n-1]}(|\,\tilde{v}^n\xi'><\tilde{v}^n\xi\,|)) \;=\; D_{n+1}D_n \dots D_1(|\,\xi'><\xi\,|).$$

□

Continuing our interpretation of Lemma 3.1 we may say that the product $D_{n+1}D_n \dots D_1$ encodes information about the defect arising from the $(n+1)$-th power of an adapted isometry. This can be made more concrete by using probabilistic language. In Section 2 we gave an interpretation of an adapted isometry as a time evolution of a stationary stochastic process. Now we can interpret the term $p_{[0,n]}\tilde{v}^{n+1}\xi$ appearing in Theorem 3.3*(b)* as the best predictor of $\tilde{v}^{n+1}\xi$, a random variable of time $n+1$, given the information available up to time n. More precisely: the best non-linear one-step predictor in the mean square sense. "Non-linear" refers to the fact that not only the linear span of random variables of the process is used for prediction but the whole algebra. A survey on several topics in prediction theory is [4]. In our setting the $(n+1)$-th prediction error $f_{n+1}(\xi)$ (for $\xi \in \mathcal{K}_0$) is given by

$$f_{n+1}(\xi) := \|\tilde{v}^{n+1}\xi - p_{[0,n]}\tilde{v}^{n+1}\xi\| = \left(\|\xi\|^2 - \|p_{[0,n]}\tilde{v}^{n+1}\xi\|^2\right)^{\frac{1}{2}}.$$

From Theorem 3.3 we get the following formula:

Corollary 3.4. *For all $\xi \in \mathcal{K}_0$ and $n \in \mathbb{N}_0$*

$$f_{n+1}(\xi)^2 + <\Omega_{n+1}, D_{n+1}D_n \dots D_1(|\,\xi><\xi\,|)\Omega_{n+1}> = \|\xi\|^2.$$

It is interesting that in linear prediction theory there is a similar formula for linear prediction errors which also involves products (of different quantities, cf. [4], Th.5.1 or [5], II.5, II.6). In that case direct sums of Hilbert spaces are used where in our theory there are tensor products. On the other hand expressions like $D_{n+1}D_n \dots D_1(|\,\xi><\xi\,|)$ are well-known in quantum theory as an irreversible time evolution converting a pure state of a quantum system into a mixed one (cf. [2]).

As a first application of this correspondence between prediction and quantum dynamical time evolutions we want to analyse asymptotics, i.e. the behaviour for large time ($n \to \infty$). For this we need some technical facts about density matrices. For the convenience of the reader we also include an elementary proof for them.

Lemma 3.5. *Consider sequences* $\{\mathcal{K}_n\}$ *of Hilbert spaces,* $\{\Omega_n\}$ *of unit vectors,* $\{\rho_n\}$ *of density matrices such that* $\Omega_n \in \mathcal{K}_n$, $\rho_n \in \mathcal{T}_1^+(\mathcal{K}_n)$ *for all* n. *The following assertions are equivalent (for* $n \to \infty$*):*

(1) $<\Omega_n, \rho_n \Omega_n> \to 1$

(2) $\| \rho_n - |\Omega_n><\Omega_n| \|_1 \to 0 \quad (\|\cdot\|_1$ *denotes trace norm)*

(3) For any sequence $\{x_n\}$ *with* $x_n \in \mathcal{B}(\mathcal{K}_n)$, $\|x_n\| = 1$ *for all* n*:*

$Tr(\rho_n x_n) - <\Omega_n, x_n \Omega_n> \to 0.$

Proof: (*2*) $\Rightarrow$ (*3*) is clear and from $<\Omega_n, \rho_n \Omega_n> = Tr(\rho_n \ |\Omega_n><\Omega_n|)$ we quickly infer (*3*) $\Rightarrow$ (*1*). It remains to prove that (*1*) $\Rightarrow$ (*2*):

Write $\rho_n = \sum_i \alpha_i^{(n)} \, |\epsilon_i^{(n)}><\epsilon_i^{(n)}|$ with $\alpha_i^{(n)} \geq 0$, $\sum_i \alpha_i^{(n)} = 1$ and $\{\epsilon_i^{(n)}\}$ an orthonormal basis of $\mathcal{K}_n$. From (*1*) we get

$$<\Omega_n, \left(\sum_i \alpha_i^{(n)} \, |\epsilon_i^{(n)}><\epsilon_i^{(n)}| \right) \Omega_n> = \sum_i \alpha_i^{(n)} \, |<\epsilon_i^{(n)}, \Omega_n>|^2 \to 1.$$

If $i = 1$ is an index with $\alpha_1^{(n)} = \max_i \alpha_i^{(n)}$ for all n then because of $\sum_i \alpha_i^{(n)} = 1 = \sum_i \ |<\epsilon_i^{(n)}, \Omega_n>|^2$ we infer $\alpha_1^{(n)} \to 1$, i.e. $\sum_{i \neq 1} \alpha_i^{(n)} \to 0$ and

$$|<\epsilon_1^{(n)}, \Omega_n>| \to 1, \text{ i.e. } \| \, |\epsilon_1^{(n)}><\epsilon_1^{(n)}| - |\Omega_n><\Omega_n| \, \|_1 \to 0.$$

Finally $\quad \|\rho_n - |\Omega_n><\Omega_n| \, \|_1 = \| \sum_i \alpha_i^{(n)} \, |\epsilon_i^{(n)}><\epsilon_i^{(n)}| - |\Omega_n><\Omega_n| \, \|_1$

$$\leq \|\alpha_1^{(n)} \, |\epsilon_1^{(n)}><\epsilon_1^{(n)}| - |\Omega_n><\Omega_n| \, \|_1 + \| \sum_{i \neq 1} \alpha_i^{(n)} \, |\epsilon_i^{(n)}><\epsilon_i^{(n)}| \, \|_1$$

$$\leq |\alpha_1^{(n)} - 1| + \| \ |\epsilon_1^{(n)}><\epsilon_1^{(n)}| - |\Omega_n><\Omega_n| \ \|_1 + \sum_{i \neq 1} \alpha_i^{(n)} \to 0.$$

□

Proposition 3.6. *Assume $\xi \in \mathcal{K}_0$, $\|\xi\| = 1$. The following assertions are equivalent:*

(1) $\lim_{n\to\infty} f_n(\xi) = 0$,

(2) $\lim_{n\to\infty} \left(D_n \dots D_1(|\,\xi><\xi\,|) - |\,\Omega_n><\Omega_n\,| \right) = 0$

(weak or with respect to the trace norm).

Proof: Take the formula for $f_n(\xi)$ in Corollary 3.4 and then apply Lemma 3.5 with $\rho_n = D_n \dots D_1(|\,\xi><\xi\,|)$. □

Note that the sequence $\{f_n(\xi)\}$ of prediction errors is in any case a non-increasing sequence of non-negative numbers and thus there is always a limit $f_\infty(\xi) := \lim_{n\to\infty} f_n(\xi)$. This is immediate because the time interval used for prediction increases and there is more and more information available. Proposition 3.6 gives a criterion for this limit to be zero, i.e. for prediction becoming perfect for $n \to \infty$. To formulate this criterion verbally we state some definitions.

Definition 3.7. *(a) A stationary stochastic process given by $\mathcal{K}_0$ and an adapted isometry $\tilde{v}$ is called deterministic with respect to the filtration $\{\mathcal{K}_n, \Omega_n\}_{n=0}^{\infty}$ if $f_\infty(\xi) := \lim_{n\to\infty} f_n(\xi) = 0$ for all $\xi \in \mathcal{K}_0$.*

(b) If the conditions of Lemma 3.5 are fulfilled then the sequence $\{|\,\Omega_n><\Omega_n\,|\}$ is called absorbing for the sequence $\{\rho_n\}$ of density matrices.

(c) Let $\{\tilde{D}_n\}$ be a sequence of maps with $\tilde{D}_n : \mathcal{T}_1^+(\mathcal{K}_0) \to \mathcal{T}_1^+(\mathcal{K}_n)$. If $\{|\,\Omega_n><\Omega_n\,|\}$ is absorbing for all sequences $\{\tilde{D}_n(\rho)\}$ with $\rho \in \mathcal{T}_1^+(\mathcal{K}_0)$ then we call $\{|\,\Omega_n><\Omega_n\,|\}$ absorbing for $\{\tilde{D}_n\}$.

(d) An adapted isometry and the associated process are called homogeneous if all Hilbert spaces $\mathcal{K}_n$ can be identified with a Hilbert space $\mathcal{K}$, all unit vectors $\Omega_n \in \mathcal{K}_n$ with a unit vector $\Omega \in \mathcal{K}$ and all unitaries $u_n \in \mathcal{B}(\mathcal{K}_{[n-1,n]})$ with a unitary $u \in \mathcal{B}(\mathcal{K} \otimes \mathcal{K})$.

Corollary 3.8. *A stationary stochastic process given by $\mathcal{K}_0$ and $\tilde{v}$ is deterministic with respect to $\{\mathcal{K}_n, \Omega_n\}$ if and only if $\{|\Omega_n\rangle\langle\Omega_n|\}$ is absorbing for $\{D_n \dots D_1\}$.*

In the homogeneous case we can further identify all operators D_m with an operator $D : \mathcal{T}(\mathcal{K}) \to \mathcal{T}(\mathcal{K})$. The criterion for determinism now tells us that $|\Omega\rangle\langle\Omega|$ should be absorbing for the semigroup $\{D^n\}_{n=0}^{\infty}$. Absorbing vector states for positive semigroups are a well-known subject in mathematics and physics (cf. [3,2]) and we make contact to it at this point. The problem can be approached via spectral theory for D and is closely related to ergodic theory.

4. Processes with finitely many values

In this section we want to illustrate the theory developed above. The emphasis is on examples, we do not try to be exhaustive.

As shown in Section 2 the concept of an adapted isometry includes commutative stationary stochastic processes. This raises the question how our results can be interpreted in this case and whether it is possible to discuss them with more traditional probabilistic means. This is indeed the case.

Take $\{1,\ldots,d\}^{\mathbb{N}_0}$ (with some natural number $d \geq 2$) with the infinite product μ of the probability measure giving equal weight to all elements of $\{1,\ldots,d\}$. Then it is easy to check that a measure preserving transformation $\tilde{\tau}$ of this probability space is adapted with respect to the natural filtration (i.e. gives rise to an adapted isometry when the construction in Section 2 is performed) if for all $\omega = \{\omega_n\}_{n=0}^{\infty} \in \{1,\ldots,d\}^{\mathbb{N}_0}$ and all $n \in \mathbb{N}_0$ the values $\{(\tilde{\tau}\omega)_i\}_{i=0}^{n}$ depend only on $\{\omega_i\}_{i=0}^{n+1}$. Analogous to the argument given in the proof of Proposition 2.2 it follows that an adapted transformation $\tilde{\tau}$ can be decomposed as an infinite product $\tilde{\tau} = \lim_{n\to\infty} \tau_n \ldots \tau_1$, where τ_n is a permutation of $\{1,\ldots,d\}^{\{0,\ldots,n\}}$ which acts identically on $\{1,\ldots,d\}^{\{0,\ldots,n-2\}}$. Note that the value of $(\tilde{\tau}\omega)_n$ is already determined by $\tau_{n+1}\tau_n \ldots \tau_1\omega$, i.e. the limit is well defined. Our simplifying assumption that $u_n \in \mathcal{B}(\mathcal{K}_{[n-1,n]})$ (see the second remark after Proposition 2.2) means here that τ_n is simply a permutation of $\{1,\ldots,d\}^{\{n-1,n\}}$.

Now the prediction problem considered in Section 3 can be formulated as a game. If we are given only $\omega_0,\ldots,\omega_n$ of some $\omega = \{\omega_n\}_{n=0}^{\infty} \in \{1,\ldots,d\}^{\mathbb{N}_0}$ then in general it is not possible to determine $(\tilde{\tau}^{n+1}\omega)_0$. We may try to guess. It depends on $\tilde{\tau}$ how much uncertainty we have to endure. Indeed the prediction errors show the amounts of errors (in the mean square sense) which are inevitable even with the best strategy. More precisely, let ξ be any (complex-valued) function on $\{1,\ldots,d\}$. Given certain values $\omega_0,\ldots,\omega_n$ there is a probability distribution $\mu_{\omega_0,\ldots,\omega_n}$ on $\{1,\ldots,d\}$ for $(\tilde{\tau}^{n+1}\omega)_0$ conditioned by these values. Elementary probability theory shows that the best prediction of $\xi((\tilde{\tau}^{n+1}\omega)_0)$ given $\omega_0, \ldots, \omega_n$ is obtained as expectation of ξ with respect to $\mu_{\omega_0,\ldots,\omega_n}$ with the variance $Var(\xi, \mu_{\omega_0,\ldots,\omega_n})$ as squared error. Then the total mean square error $f_{n+1}(\xi)$ is obtained by averaging over all possible $\omega_0, \ldots, \omega_n$: $f_{n+1}(\xi)^2 = \frac{1}{d^{n+1}} \sum_{\omega_0,\ldots,\omega_n} Var(\xi, \mu_{\omega_0,\ldots,\omega_n})$. This justifies the interpretation given above.

In Corollary 3.4 we derived an alternative expression in terms of a product of completely positive maps $D_m : M_d \to M_d$. Here M_d denotes the $d \times d$-matrices $(= \mathcal{T}(\mathbb{C}^d))$. We have

$$f_{n+1}(\xi)^2 + < \Omega_{n+1}, D_{n+1} D_n \ldots D_1(|\xi><\xi|) \Omega_{n+1} >= \|\xi\|^2 .$$

Here $\Omega_{n+1} = (1, 1, \ldots, 1) \in \mathbb{C}^d$. We want to write down the operators D_m more explicitely. For this consideration we drop the index. The operator D is derived from a permutation τ of $\{1, \ldots, d\}^2$ giving rise to an isometry $v : \mathbb{C}^d \to \mathbb{C}^d \otimes \mathbb{C}^d$ so that for $\rho \in M_d$ we have $D(\rho) = Tr_{\mathbb{C}^d}(v \rho v^*)$ (with trace evaluated on the left). We want to calculate coordinates with respect to the normalized canonical basis $\{|i>\}_{i=1}^d$ of $\mathbb{C}^d$, i.e. $|i>$ has entry $\sqrt{d}$ at position i and zero elsewhere. Let us write $i \xrightarrow{k} j$ if the first component of $\tau(i, k)$ is j. Then a straightforward computation yields

Lemma 4.1. $$v \, |j> = \frac{1}{\sqrt{d}} \sum_{i \xrightarrow{k} j} |i> \otimes |k>$$

$$D_{kl,ij} := < k, D(|i><j|) l > = \frac{1}{d} \sharp \{ r : r \xrightarrow{k} i \ \ and \ \ r \xrightarrow{l} j \},$$

where $\sharp$ counts the number of elements.

Some observations about these coordinates of D are immediate: There is a symmetry $D_{kl,ij} = D_{lk,ji}$. Further, fixing k, l and summing over i, j

always yields one, which proves the surprising fact that D with respect to the normalized canonical basis gives rise to a (row-)stochastic $d^2 \times d^2$-matrix. Its entries are a kind of transition probabilities for pairs when applying τ, refining the transition probabilities for individuals which are included as $D_{kk,ii} = \sharp\{r : r \stackrel{k}{\longrightarrow} i\}$.

Putting all this together we have proved the following combinatorial formula which summarizes the computation of prediction errors in this setting:

Proposition 4.2. *For all $n \in \mathbb{N}_0$, $i \in \{1, \ldots, d\}$:*

$$f_{n+1}(|i\rangle)^2 + \frac{1}{d} \sum_{k,l=1}^{d} (D_{n+1} D_n \ldots D_1)_{kl,ii} = 1.$$

The sum is a column sum of a (row-)stochastic $d^2 \times d^2$-matrix which is given as the product of the (row-)stochastic $d^2 \times d^2$-matrices associated to the operators D_m as in Lemma 4.1.

Of course the occurrence of stochastic matrices simplifies the asymptotic theory. See [7] for some basic facts about stochastic matrices, in particular: A set of indices is called essential for a stochastic matrix if by successive transitions allowed by the matrix it is possible to go from any element of the set to any other, but it is not possible to leave the set. An index not contained in an essential set is called inessential.

Proposition 4.3. *For the processes considered in this section the following assertions are equivalent:*

(1) The process is deterministic.

(2) All entries of the stochastic matrices associated to the products $D_n \dots D_1$ which do not belong to an ii-column ($i \in \{1, \dots, d\}$) tend to zero for $n \to \infty$.

And if the process is homogeneous ($D_m \simeq D$ for all m):

(3) Indices ij with $i \neq j$ are inessential for the stochastic matrix associated to D.

Proof: Determinism means that $f_n(|\, i >) \to 0$ for all $i \in \{1, \dots, d\}$ and $n \to \infty$. By Proposition 4.2 this is the case if and only if the column sums $\sum_{k,l=1}^{d} (D_{n+1} D_n \dots D_1)_{kl,ii}$ tend to d. But the sum of all entries is d^2 and none of the column sums can exceed d (obvious from Proposition 4.2). This proves *(1)* $\Leftrightarrow$ *(2)*. It is a general fact that for powers of a single stochastic matrix D we have the equivalence *(2)* $\Leftrightarrow$ *(3)* (cf. [7], chapter 4). □

Especially condition *(3)* of Proposition 4.3 is very easy to check, at least for matrices of moderate size. We give an example: Choose $d = 3$ and consider the homogeneous process generated by the permutation τ of $\{1, 2, 3\}^2$ given by the cycle $(11, 12, 13, 23, 22, 21, 31, 32, 33)$. Using Lemma 4.1 we can compute the associated stochastic matrix. The result (with indices ordered as follows: $11, 22, 33, 12, 21, 13, 31, 23, 32$) is shown on the next page. For example the non-zero entries in the fourth row (with index 12) are obtained from $1 \xrightarrow{1} 1$, $1 \xrightarrow{2} 1$ and $3 \xrightarrow{1} 3$, $3 \xrightarrow{2} 3$ and $2 \xrightarrow{1} 3$, $2 \xrightarrow{2} 2$. It is easy to check that starting from any index ij we can in at most two steps reach the essential set $\{11, 22, 33\}$. With Proposition 4.3*(3)* it follows that the process is deterministic.

$$\frac{1}{3}\begin{pmatrix} 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

References

1. L. Accardi, A. Frigerio, J. Lewis: Quantum Stochastic Processes, Publ.RIMS Kyoto **18**, 94-133 (1982)
2. R. Alicki, K. Lendi: *Quantum Dynamical Semigroups and Applications*, LNP 286, Springer-Verlag (1987)
3. W. Arveson: Pure E_0-semigroups and absorbing states, Commun. Math. Phys. **187**, No.1, 19-43 (1997)
4. R. S. Bucy: *Lectures on Discrete Time Filtering*, Springer-Verlag (1994)
5. C. Foias, A.E. Frazho: *The Commutant Lifting Approach to Interpolation Problems*, Birkhäuser-Verlag (1990)
6. R. Kadison, J. Ringrose: *Fundamentals of the Theory of Operator Algebras, Vol. I, II*, Academic Press (1986)
7. E. Seneta: *Non-negative Matrices and Markov Chains*, Springer-Verlag (1980)