Discrete Quantum Markov Chains

Ulrich Faigle\textsuperscript{1} and Alexander Schönhuth\textsuperscript{2}

\textsuperscript{1} Mathematical Institute
Center for Applied Computer Science
University of Cologne
Weyertal 80, 50931 Köln, Germany
faigle@zpr.uni-koeln.de

\textsuperscript{2} Department of Mathematics
University of California, Berkeley
970 Evans Hall, Berkeley, CA 94720-3840, USA
alexsch@math.berkeley.edu

Abstract. A framework for finite-dimensional quantum Markov chains on Hilbert spaces is introduced. Quantum Markov chains generalize both classical Markov chains with possibly hidden states and existing models of quantum walks on finite graphs. Quantum Markov chains are based on Markov operations that may be applied to quantum systems and include quantum measurements, for example. It is proved that quantum Markov chains are asymptotically stationary and hence possess ergodic and entropic properties. With a quantum Markov chain one may associate a quantum Markov process, which is a stochastic process in the classical sense. Generalized Markov chains allow a representation with respect to a generalized Markov source model with definite (but possibly hidden) states relative to which observables give rise to classical stochastic processes. It is demonstrated that this model allows for observables to violate Bell’s inequality.

Keywords. Asymptotic stationary, Bell’s inequality, hidden state, Markov source, quantum computation, quantum Markov chain, quantum statistics, quantum walk, stochastic process

1 Introduction

In mathematical models of physical situations, states are usually represented by probability distributions on certain basic configurations. This idea has also proved powerful in data analysis and communication where the model of a Markov source (with possibly "hidden states" as basic configurations) is often successfully employed in order to extract information from time series (see, e.g., Choi \textit{et al.} \cite{6} or Elliot \textit{et al.} \cite{9}).

The mathematical model of quantum mechanics generalizes the classical model in that it represents states by density matrices (\textit{i.e.}, complex matrices $P$ of the form $P = A^*A$ with trace $\text{tr}(P) = 1$) on quantum systems, whose descriptions do not appear to be so clearly based on hidden states (\textit{cf.} Einstein \textit{et al.} \cite{8}). The quantum model is not restricted to physics. It has also been used to formalize an extension of the computational model of a (probabilistic) Turing machine. This extension has celebrated a
striking success in Shor’s polynomial algorithm for integer factorization and has stimulated research into quantum computation as well as quantum information theory (see, e.g., Nielsen and Chuang). At the same time, a theory of the statistics of quantum measurements and operations is emerging (see Barndorff-Nielsen et al.). Moreover, a model for quantum walks on graphs has been proposed by Aharonov et al. One might expect that a full theory of quantum processes will increase the power of the model of stochastic processes to a similar degree.

The present article introduces a framework for discrete quantum Markov chains on finite-dimensional Hilbert spaces in the spirit of classical Markov sources (with possibly hidden states), which appear in this model as those quantum Markov chains that allow state representations by diagonal density matrices. The quantum walks of Aharonov et al. turn out to be special cases of quantum Markov chains. Moreover, as in the classical case, repeated measurements and operations on systems yield Markov chains and produce observable values that give rise to quantum Markov processes. We prove that quantum Markov chains are asymptotically stationary, which means that they have ergodic and entropic properties and hence appear quite promising for statistical and information-theoretic analysis.

We show that quantum Markov processes are (classical) stochastic processes with a finite dimension in the sense of Ito et al. (see also Jaeger). Therefore, quantum Markov processes can be represented by state vectors (rather than density matrices) in finite-dimensional real coordinate spaces. We provide a representation that suggests to interpret general Markov processes as sequences of signals which emanate–as in the classical case–from a source with a finite number of hidden states. The signal seen depends just on the hidden state of the source. However, we must make a concession to this general model: transitions between hidden states may no longer be described by exclusively nonnegative parameters.

It has long been realized that "negative probabilities" provide a mathematically very reasonable approach to quantum mechanical modeling (see, e.g., Dirac, Feynman or Khrennikov). However, negative probabilities are counterintuitive and, therefore, not suited for models where "probabilities" are to reflect degrees of subjective belief. One would like to evaluate observed events by nonnegative probabilities. Our model of generalized Markov sources with definite (but hidden) states achieves the latter goal. The observable stochastic processes produced by such sources fit into the classical probability model of Kolmogorov – even though they may intrinsically depend on state descriptions with negative parameters. (The situation is thus quite similar to the widely accepted quantum state description in terms of density matrices, which may very well contain negative coefficients.) We show in the last section that this model of definite states allows for Bell’s inequality to be violated by classically observed variables.

2 States

Let \( V \) be a vector space over the field \( \mathbb{C} \) of complex numbers of finite dimension \( N \) and fix a basis \( E = \{ e_1, \ldots, e_N \} \) for reference. Using the Dirac notation, we denote
by \( |v\rangle \in \mathbb{C}^N \) the coordinate (column) vector of \( v \in V \) relative to the standard basis \( E \) and denote by \( \langle v | = (|v\rangle)^* \) its conjugated transpose. \( V \) becomes a Hilbert space \( \mathcal{H} \) with respect to the inner product

\[
\langle v | w \rangle := \langle v | I | w \rangle = \sum_{i=1}^{N} x_i y_i
\]

for \( v = \sum_{i=1}^{N} x_i e_i \), \( w = \sum_{i=1}^{N} y_i e_i \), where \( I \) is the \((N \times N)\)-identity matrix. We are interested in self-adjoint (a.k.a. Hermitian) \((N \times N)\)-matrices which, by definition, are equal to their conjugate transpose:

\[
Q = [q_{ij}] = [\overline{q_{ji}}] = Q^* .
\]

A self-adjoint matrix \( Q = [q_{ij}] \) has real eigenvalues \( \lambda_i \) and a corresponding orthonormal set \( \{v_1, \ldots, v_N\} \) of eigenvectors, yielding the spectral representation

\[
Q = \sum_{i=1}^{N} \lambda_i |v_i\rangle \langle v_i| \quad \text{and hence the trace } \quad \text{tr}Q = \sum_{i=1}^{N} q_{ii} = \sum_{i=1}^{N} \lambda_i .
\]

The self-adjoint matrix \( Q \) is said to be nonnegative if

\[
\langle v | Q | v \rangle \geq 0 \quad \text{for all } v \in \mathcal{H}.
\]

The spectral representation shows that nonnegative matrices are characterized by the fact that all eigenvalues are nonnegative real scalars. In accord with the terminology of quantum mechanics, we refer to a nonnegative matrix \( Q \) with trace \( \text{tr}Q = 1 \) as a density matrix or (quantum) state.

Not only the eigenvalues but also all diagonal elements of a nonnegative matrix are nonnegative real numbers. Hence the diagonal elements of an arbitrary quantum state \( Q \) specify a probability distribution on the standard basis \( E \).

### 2.1 The state space

Let \( S \) denote the collection of all self-adjoint matrices. \( S \) can be interpreted as a real vector space as follows. Any complex \((N \times N)\)-matrix \( Q \) admits a unique representation of the form

\[
Q = A + iB \quad \text{with } A, B \in \mathbb{R}^{N \times N} .
\]

Thus \( Q \) is self-adjoint if and only if \( A \) is symmetric and \( B \) skew-symmetric:

\[
Q^* = Q \iff A^T = A \quad \text{and } B^T = -B .
\]

So we may view \( S \) as the real vector space of all pairs \((A, B)\), where \( A \) is a real symmetric and \( B \) a real skew-symmetric matrix. Moreover, since all diagonal elements of a real skew-symmetric matrix are 0, we find for each \( Q = A + iB \in S \):

\[
\text{tr}Q = \text{tr}A + i \cdot \text{tr}B = \text{tr}A .
\]
Let $\mathcal{S}_+ \subseteq \mathcal{S}$ denote the collection of nonnegative matrices. It is clear that $\mathcal{S}_+$ is closed under taking linear combinations with (real) nonnegative scalar. So $\mathcal{S}_+$ is a convex cone in $\mathcal{S}$, which contains the null-matrix $0$. In fact, we find for any nonnegative matrix $Q \in \mathcal{S}_+$:
\[
\text{tr}Q = 0 \iff Q = 0.
\]
Hence each nonnegative matrix $Q \in \mathcal{S}_+$ can be presented in the form
\[
Q = \text{tr}(Q)Q' \quad \text{where } Q' = 0 \text{ or } Q' \in \mathcal{S}_+ \text{ is a density matrix.}
\]

## 3 Quantum Markov processes

In order to motivate our model for quantum Markov chains, we first recall the model of a classical (hidden) Markov source with finite alphabet $\Sigma$ and then introduce the model of quantum Markov chains as a natural generalization.

### 3.1 Classical Markov sources

One considers a source that is assumed to be always in one of $N$ possible ”hidden states” $i \in \Omega$, which cannot be observed directly. However, there is a fixed information function $X : \Omega \to \Sigma$ into some alphabet $\Sigma$ and the source emits the symbol $X(i) \in \Sigma$ whenever it is in the hidden state $i \in \Omega$. The source works as follows:

- At time $t = 0$ the empty message $X_0 = \square$ is sent. According to a given probability distribution $p^0$ on $\Omega$, the source then enters a hidden state $j$ and produces $X_1 = X(j)$. Then the source changes from $j$ to a hidden state $i$ with probability $m_{ij}$ and produces the next symbol $X_2 = X(j)$ etc.

The corresponding transition matrix $M = [m_{ij}]$ has the transition probability distributions as column vectors:
\[
m_{ij} \geq 0 \quad \text{and} \quad \sum_{i=1}^{N} m_{ij} = 1 \quad (j = 1, \ldots, N).
\]
For all $a \in \Sigma$, we define the matrices $S^a = [s^a_{ij}]$ via
\[
s^a_{ij} = \begin{cases} m_{ij} & \text{when } X(i) = a \\ 0 & \text{when } X(i) \neq a. \end{cases}
\]
As usual, we denote by $\Sigma^* = \bigcup_{t \geq 0} \Sigma^t$ the collection of all finite words with letters from $\Sigma$. $\Sigma^*$ is a semigroup under the concatenation $vw$ of words $v, w \in \Sigma^*$ with the empty word $\square$ as neutral element. The concatenation operation suggests to define matrices $S^w$ for all words $w = w_1w_2 \ldots w_t$ via the matrix product
\[
S^w = S^{w_1}S^{w_2} \cdots S^{w_t} \quad \text{and} \quad S^\square = S^0 = I \quad (\text{identity matrix}).
\]
It is not difficult to see that \( p'_w = S^w p^0 \) is a coordinate vector whose component \( p'_w(i) \) is exactly the probability for the source to be in the hidden state \( i \) if the word \( w \) (i.e., the event \( \{ X_1 = w_1, \ldots, X_t = w_t \} \)) has been observed, which implies

\[
p(w) = \Pr\{ X_1 = w_1, \ldots, X_t = w_t \} = \sum_{i=1}^{N} p'_w(i) = 1^T S^w p^0,
\]

where \( 1^T = [1,1,\ldots,1] \). Hence, if \( p(w) \neq 0 \), we can express \( S^w p^0 \) in the form

\[
S^w p^0 = p(w)p_w, \text{ where } p_w \text{ is a probability distribution on } \Omega.
\]

In the case \( p(w) \neq 0 \), the conditional probabilities are given by

\[
\Pr\{ X_{t+1} = a | w \} = p(a | w) = \frac{p(aw)}{p(w)} = 1^T S^a p_w,
\]

which means \( S^a p_w = p(a | w)p_{wa} \). So \( (X_t) \) is a stochastic process whose \( t \text{th ground state} \) we define as the probability distribution

\[
p_t^0 = M_t^0 p^0 = (\sum_{a \in \Sigma} S^a)^t p^0 = \left( \sum_{w \in \Sigma^t} S^w \right) p^0 = \sum_{w \in \Sigma^t} p(w) p_w .
\]

The Markov source is stationary if \( M p^0 = p^0 \), which means that \( p^0 \) is an eigenvector of \( M \) with eigenvalue \( \lambda = 1 \).

**Sampling.** Let \( v = v_1 \ldots v_m \in \Sigma^* \) be a given word. Then the probability for the Markov source to produce \( v \) in the time periods \( t+1, \ldots, t+m \) is encoded in the vector

\[
S^v p^t \text{ with the property } \Pr\{ X_{t+1} = v_1, \ldots, X_{t+m} = v_m \} = 1^T S^v p^t.
\]

Sampling the Markov source for \( v \), one would like the expected values of the sampling averages to converge in the limit, i.e., one would like

\[
\overline{p}(v) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} 1^T S^w p^k
\]

to exist. One can prove (see Faigle and Schönhuth [10], for example) that \( \overline{p}(v) \) always exists. In fact, the existence of

\[
\overline{p} = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} p^0
\]

is guaranteed, which yields \( \overline{p}(v) = 1^T S^v \overline{p} \). \( \overline{p} \) is a probability distribution on \( \Omega \) and an eigenvector of \( M \). Hence Markov sources are asymptotically stationary.

**Remark.** It is known for a stochastic process \( (X_t) \) with finite alphabet \( \Sigma \) that the property of being asymptotically stationary is equivalent with \( (X_t) \) having general ergodic properties. In particular, \( (X_t) \) then possesses a well-defined entropy rate

\[
H_\infty = \lim_{t \to \infty} H(X_t | X_0 \ldots X_{t-1}) = \lim_{t \to \infty} \frac{1}{t} H(X_0, \ldots, X_{t-1}).
\]

(See, e.g., Gray [13,14] for details).
3.2 Quantum Markov chains

Let $\mathcal{H}$ be an $N$-dimensional Hilbert space with state space $\mathcal{S}$. It is convenient to fix a subspace $\mathcal{V} \subseteq \mathcal{S}$ and give the definition relative to $\mathcal{V}$. A quantum Markov operation is now specified by a set $\mathcal{M} = \{\sigma^a\}$ of linear operators $\sigma^a : \mathcal{V} \to \mathcal{V}$, indexed with labels $a$ from a finite alphabet $\Sigma$, such that for all $Q \in \mathcal{V}$,

(i) $\sigma^a(Q)$ is nonnegative if $Q$ is nonnegative;
(ii) $\sum_{a \in \Sigma} \text{tr}(\sigma^a(Q)) = \text{tr}(Q)$.

The Markov operation $\mathcal{M}$ acts on a quantum system in the quantum state $P \in \mathcal{V}$ according to the following probability rule:

- Let $P_a \in \mathcal{V}$ be such that $\sigma^a(P) = p(a|P)P_a$, where $p(a|P) = \text{tr}(\sigma^a(P))$.
- Then the system changes its state from $P$ to $P_a$ with probability $p(a|P)$.

Notice that (i) and (ii) ensure that the real numbers $p(a|P)$ yield a probability distribution on $\Sigma$ if $P$ is a quantum state. Moreover, $P_a$ must be a quantum state if $p(a|P) \neq 0$. The expected state after the application of $\mathcal{M}$ is the superposition

$$\psi(P) = \sum_{a \in \Sigma} p(a|P)P_a = \sum_{a \in \Sigma} \sigma^a(P) \in \mathcal{V}.$$  

The linear operator $\psi = \sum_{a \in \Sigma} \sigma^a$ is the evolution operator associated with the Markov operation $\mathcal{M}$. By property (ii) above, $\psi$ is trace-preserving and maps quantum states of $\mathcal{V}$ to quantum states of $\mathcal{V}$.

Assume that a quantum system is initially in the state $\chi_0 = P_\square \in \mathcal{V}$. Then repeated application of the Markov operation $\mathcal{M}$ yields a (quantum) state-valued random process $(\chi_t)$ such that

$$p(w) = \Pr\{\chi_t = P_w\} = \text{tr}(\sigma^w(P_\square)) \quad \text{for all } w \in \Sigma^t.$$  

We call $(\chi_t)$ a quantum Markov chain. The states $P_t = \psi^t(P_\square)$ are the ground states of the Markov chain and admit a representation as the superposition

$$P_t = \psi^t(P^0) = \left(\sum_{a \in \Sigma} \sigma^a\right)^t P_\square = \left(\sum_{w \in \Sigma^t} \sigma^w\right)P_\square = \sum_{w \in \Sigma^t} p(w)P_w.$$  

We illustrate the concept of quantum Markov chains with some examples.

**Classical Markov sources.** Let $M$ be the transition matrix of a Markov source with alphabet $\Sigma$. Representing arbitrary coordinate vectors by diagonal matrices,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N \quad \longleftrightarrow \quad \text{diag } x = \begin{bmatrix} x_1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & x_N \end{bmatrix} \in \mathcal{S},$$  

则有：

$$\sigma^a(Q) = p(a|Q)Q$$

其中 $p(a|Q) = \text{tr}(\sigma^a(Q))$.
we let $\mathcal{V} \subseteq \mathcal{S}$ be the subspace of all diagonal matrices. The map
\[
\text{diag } x \rightarrow \sigma^a(\text{diag } x) = \text{diag } S^a x
\]
is linear on $\mathcal{V}$. In view of $\text{tr}(\text{diag } x) = 1^T x$, it is clear that $\mathcal{M} = (\sigma^a)$ is a quantum Markov operation relative to $\mathcal{V}$.

**Quantum measurements.** A measurement with finite scale $\Sigma$ on an $N$-dimensional quantum system consists of a collection of nonnegative matrices $M^a \in \mathcal{S}_+$, indexed by the possible outcomes $a \in \Sigma$, that sum to the identity:
\[
\sum_{a \in \Sigma} M^a = I.
\]
If the quantum system is in the state $Q \in \mathcal{S}_+$, the number $p(a|Q) = \text{tr}(M^a Q)$ is the probability for the measurement to result in the outcome $a$. Since $I$ is trivially trace-preserving, the measurement is a quantum Markov operation $\mathcal{M} = (\sigma^a)$, where $\sigma^a(Q) = M^a Q$.

**Quantum operations.** An operation $\mathcal{O}$ on an $N$-dimensional quantum system is usually specified by a finite number of $(N \times N)$-matrices $O_i$ so that
\[
\sum_i O_i^* O_i = I.
\]
The result of the operation $\mathcal{O}$ is the transformation of the system from state $Q \in \mathcal{S}_+$ into the state
\[
\mathcal{O}(Q) = \sum_i O_i^* Q O_i.
\]
It is not difficult to verify that $\mathcal{O}(Q)$ is indeed a density matrix. So $\mathcal{M} = \{\mathcal{O}\}$ is a quantum Markov operation (with a unary alphabet $\Sigma$).

**Evolution of wave functions.** A vector $v$ with $\langle v|v \rangle = 1$ in the $N$-dimensional Hilbert space $\mathcal{H}$ is considered to be wave function and defines the quantum state
\[
P^0 = |v\rangle \langle v| \quad \text{with} \quad P^0_{ii} = |x_i|^2 \quad \text{for any} \quad |v\rangle = \sum_{i=1}^N x_i |e_i\rangle \in \mathbb{C}^N.
\]
If $U$ is a unitary matrix (i.e, $U^* U = I$ holds), then also $Uv$ is a wave function, giving rise to the state
\[
P^1 = |Uv\rangle \langle Uv| = U^* |v\rangle \langle v| U = U^* P^0 U.
\]
So the linear operator $P \rightarrow U^* PU$ defines a quantum Markov operation on $\mathcal{V} = \mathcal{S}$ with Markov chain $(P^t)$. Let $\{u_1, \ldots, u_N\}$ be an orthonormal basis of eigenvectors of $U$ and assume
\[
U = \sum_{i=1}^N \lambda_i |u_i\rangle \langle u_i| \quad \text{and} \quad |v\rangle = \sum_{i=1}^N y_i |u_i\rangle.
\]
Then we have
\[ U^t = \sum_{i=1}^{N} \lambda_i^t |u_i\rangle\langle u_i| \quad \text{and} \quad P^0 = |v\rangle\langle v| = \sum_{i=1}^{N} |y_i|^2 |u_i\rangle\langle u_i|. \]

Consequently, the states of the Markov chain \((P^t)\) are
\[ P^t = (U^t)^* P^0 U^t = \sum_{i=1}^{N} |\lambda_i|^t |y_i|^2 |u_i\rangle\langle u_i|. \quad (1) \]

**Quantum walks on graphs.** A model for a quantum walk on a graph with a finite set \(V\) of nodes has been proposed by Aharonov et al. [1] (see also Kempe [17]) for regular graphs \(G\) relative to a wave function evolution on the set of edges. A natural generalization of that model is the following.

Given an arbitrary graph \(G = (V, A)\) with a set \(V\) of nodes and a set \(A \subseteq V \times V\) of \(N = |A|\) (directed) edges, we consider the \(N\)-dimensional Hilbert space \(\mathcal{H}_A\) whose standard basis corresponds to \(A\). The quantum walk on \(G\) is now described in terms of a quantum Markov operation with a quantum state-preserving map \(\mu\). The walk on \(G\) starts in an initial quantum state \(P^0\) and consists of changing states according to
\[ P^t = \mu(P^{t-1}) \quad \text{for all} \quad t = 1, 2, \ldots \]

The probability for the quantum walk to be in node \(v \in V\) at time \(t\) is defined to be
\[ p_t(v|P^0) = \sum_{w:(v,w) \in A} \langle v,w|P^t|v,w\rangle, \]
where \(\langle v,w|P^t|v,w\rangle = P^t_{(v,w),(v,w)}\) is the evaluation of \(P^t\) at position \((v,w) \in A\) on the diagonal.

Aharonov et al. [1] assume a wave function \(v\) and a unitary matrix \(U\) and consider the quantum Markov chain
\[ P^t = \mu(P^{t-1}) = U^* P^{t-1} U \quad \text{with} \quad P^0 = |v\rangle\langle v|. \]

Making use of the state representation [1], they derive an ergodic property of their quantum walk and show that the following limit of averages exists in their model:
\[ \overline{p}(v, P^0) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} p_k(v|P^0). \]

We will see in Theorem [Section 4] that all quantum Markov chains (and hence all unitary evolutions and quantum walks) are asymptotically stationary, which implies much stronger ergodic properties.
3.3 Quantum Markov sources

The model of a quantum Markov chain \((\chi_t)\) describes the state evolution of a quantum system under the action of a quantum Markov operation \(M = \{\sigma^a\}\). Often, however, this action is accompanied by a process \((X_t)\) that can be observed as in the case of classical Markov chains or quantum measurements, for example. We model the latter situation with a source whose states are described by the quantum Markov chain. In addition, the source emits symbols \(X_t\) from the finite alphabet \(\Sigma\) as follows:

- If the source is in the quantum state \(P\) at time \(t\), it changes its state to \(P_a\) with probability \(p(a|P)\) and emits the symbol \(X_{t+1} = a \in \Sigma\) at time \(t + 1\).

So the quantum Markov chain \((\chi_t)\) gives rise to the random process \((X_t)\) with alphabet \(\Sigma\) and the probabilities

\[
\Pr\{X_1 = w_1, \ldots, X_t = w_t\} = p(w) = \text{tr} \sigma^w P_p \quad \text{for all } w = w_1 \ldots w_t \in \Sigma^*.
\]

The conditional probabilities of \((X_t)\) are given by

\[
p(v|w) = \text{tr} \sigma^v P_w \quad \text{for all } v, w \in \Sigma^* \text{ with } p(w) \neq 0.
\]

3.4 Generalized Markov processes

We call a stochastic process \((X_t)\) with finite alphabet \(\Sigma\) a generalized (quantum) Markov process if there is a finite-dimensional Hilbert space \(\mathcal{H}\) with state space \(S\) so that we can associate with each word \(w \in \Sigma^*\) some self-adjoint (but not necessarily nonnegative) matrix \(\Pi_w \in S\) with the following property:

- \(\sigma^a(\Pi_w) = p(a|w)\Pi_{wa}\) and \(\text{tr} \sigma^a(\Pi_w) = p(a|w)\).

It is not difficult to see that the definition implies

\[
\sigma^v(\Pi_w) = p(v|w)\Pi_{vw} \quad \text{and} \quad p(v|w) = \text{tr} (\sigma^v \Pi_w) \quad \text{for all } v \in \Sigma^*.
\]

As before, we refer to \(\Pi^0 = \Pi_p\) as the initial (generalized) ground state and call the linear operator \(\psi = \sum_{a \in \Sigma} \sigma^a\) the evolution operator of \((X_t)\).

**Lemma 1.** Let \((X_t)\) be a generalized Markov process with initial ground state \(\Pi^0\) and evolution operator \(\psi\). Then all ground states \(\Pi^t = \psi^t \Pi^0\) satisfy the trace condition \(\text{tr}(\Pi^t) = 1\).

\(\diamondsuit\)
4 Asymptotic stationarity

Let \( M = \{ \sigma^a | a \in \Sigma \} \) be a quantum Markov operation with respect to the finite-dimensional Hilbert space \( \mathcal{H} \) and the subspace \( V \subseteq \mathcal{S} \) with evolution operator \( \psi = \sum_{a \in \Sigma} \sigma^a \) as in Section 3.2. The main result in this section is the observation that \( M \) is asymptotically stationary in the following sense:

**Theorem 1.** For every (nonnegative) quantum state \( Q \in V \), the limit

\[
\overline{Q} = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} \psi^k(Q)
\]

exists. Moreover, \( \overline{Q} \) is an eigenstate of \( \psi \) of \( M \) with eigenvalue \( \lambda = 1 \). Consequently, also the following limit exists for all quantum states \( Q \in V \) and words \( v \in \Sigma^* \):

\[
\sigma^v(Q) = \sigma^v(\overline{Q}) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} \sigma^v \psi^k(Q).
\]

It is convenient to base the proof of Theorem 1 on the following fact.

**Lemma 2 ([10]).** Let \( V \) be a finite-dimensional normed vector space over \( \mathbb{C} \) and consider the linear operator \( F : V \to V \). The following statements are equivalent:

(a) \( \overline{v} = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} F^k(v) \) exists for all \( v \in V \).

(b) For every \( v \in V \), there exist some \( c \in \mathbb{R} \) such that \( \| F^t(v) \| \leq c \) holds for all \( t \geq 0 \).

Moreover, if (a) or (b) holds, either \( \overline{v} = 0 \in V \) or \( \overline{v} \) is an eigenvector of \( F \) with eigenvalue \( \lambda = 1 \).

We apply Lemma 2 to the subspace \( V \subseteq \mathbb{C}^{N \times N} \) that is generated by the quantum states \( P \in V \). We equip the vector space \( \mathbb{C}^{N \times N} \) with the spectral norm

\[
\| A \|_2 = \max_{\langle x | x \rangle = 1} \sqrt{\langle Ax | Ax \rangle}
\]

and observe that \( \| P \|_2 \leq 1 \) holds for all quantum states \( P \). Let us fix a basis \( \{ P_1, \ldots, P_m \} \) of quantum states \( P_i \) for \( V \) and define the linear operator \( F : V \to V \) via

\[
F\left( \sum_{i=1}^{m} r_i P_i \right) = \sum_{i=1}^{m} r_i \psi(P_i).
\]

Notice that \( F^t(P_i) = \psi^t(P_i) \) holds because \( \psi \) is linear on \( V \) and maps quantum states into quantum states. So we have \( \| F^t(P_i) \|_2 = \| \psi^t(P_i) \|_2 \leq 1 \) for all \( t \) and, therefore, by the triangle inequality

\[
\| F^t \left( \sum_{i=1}^{m} r_i P_i \right) \|_2 \leq \sum_{i=1}^{m} |r_i| \leq \sum_{i=1}^{m} |r_i|.
\]
In other words, the linear operator $F$ satisfies condition (b) and hence also (a) of Lemma 2, which proves Theorem 1.

We illustrate Theorem 1 with an application to the quantum walk model of the graph $G = (V, A)$ (see Section 3.2). We want to “sample” the quantum walk at time $t$ for the node $v \in V$. This is done by sampling all edges of the form $(v, w)$. Each such edge $(v, w)$ yields a positive sampling result with probability $P_t^{(v, w)}$, $(v, w)$ at time $t$.

Theorem 1 guarantees the existence of $\lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} P^k$ and hence the existence of $\lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} \sum_{w: (v, w) \in A} P^k_{(v, w), (v, w)}$.

Evolution of wave functions. Note that the wave functions in a wave function evolution will generally not converge. Let for example $U$ be an arbitrary unitary matrix and assume that $U$ does not have 1 as an eigenvalue. Lemma 2 then implies for any evolution $U^t \psi$:

$$\lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} U^k \psi = 0 \in \mathcal{H}.$$ 

In contrast to this null-convergence of the wave function, however, Theorem 1 says that the corresponding ground states $P^{(t)}$ converge to a quantum state:

$$\mathcal{P} = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} P^{(k)} \quad\text{and}\quad \text{tr}(\mathcal{P}) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} \text{tr}(P^{(k)}) = 1.$$ 

As matter of fact, unitary Markov chains exhibit an even stronger convergence property (see formula (2) below) as they are nonnegative in the sense of Section 4.1.

4.1 Nonnegative Markov sources

Recall that the set $\mathbb{C}^{N \times N}$ of square complex matrices forms a Hilbert space with respect to the inner product

$$\langle A, B \rangle = \text{tr}(A^* B) \quad\text{for all } A, B \in \mathbb{C}^{N \times N},$$

which induces the inner product $\langle P, Q \rangle = \text{tr}(PQ)$ on the $N^2$-dimensional (real) vector space $\mathcal{S}$ of self-adjoint matrices $P, Q$. We call a Markov source nonnegative if its evolution operator $\psi$ is a nonnegative linear operator on $\mathcal{S}$, i.e.,

$$\langle Q, \psi(Q) \rangle = \text{tr}(Q \psi(Q)) \geq 0 \quad\text{for all } Q \in \mathcal{S}.$$

So $\mathcal{S}$ admits an orthonormal basis $\{B_1, \ldots, B_{N^2}\}$ of eigenmatrices $B_j \in \mathcal{S}$ of $\psi$ with eigenvalues $\lambda_j$.

Lemma 3. $0 \leq \lambda_j \leq 1$ holds for each eigenvalue $\lambda_j$ of the nonnegative evolution operator $\psi$. 
Proof. Since $\psi$ is nonnegative, all eigenvalues are nonnegative real numbers. By Theorem 1, $\lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} \psi^k(P)$ exists for all nonnegative matrices $P \in S$ and thus, by the linearity of $\psi$, for all $Q \in S$. So Lemma 2 implies, for all $t \geq 0$,

$$\lambda_t^j = \lambda_j^j \|B_j\| = \|\psi^t(B_j)\| \leq c \quad \text{and hence} \quad \lambda_j \leq 1 .$$

Assume $\lambda_1 = \ldots \lambda_h = 1$ and $0 \leq \lambda_j < 1$ for all $j \geq h + 1$. Any quantum state $P$ can be written in the form

$$P = \sum_{j=1}^{N^2} \alpha_j B_j = B + \sum_{j=h+1}^{N^2} \alpha_j B_j \quad \text{with} \quad B = \sum_{j=1}^{h} \alpha_j B_j .$$

Hence we find

$$\psi^t(P) = B + \sum_{j=h+1}^{N^2} \alpha_j \lambda_j^j B_j \quad \text{and}$$

$$\overline{\psi^t}(P) = \frac{1}{t} \sum_{k=0}^{t-1} \psi^k(P) = B + \sum_{j=h+1}^{N^2} \alpha_j \frac{(1 - \lambda_j^j)}{t(1 - \lambda_j)} B_j .$$

Because $\lambda_j^j \to 0$ if $j \geq h + 1$, we can give be a stronger convergence guarantee than the one in Theorem 1

$$P = \lim_{t \to \infty} \overline{\psi^t}(P) = B = \lim_{t \to \infty} \psi^t(P) . \quad (2)$$

Unitary Markov sources. Let us call a Markov source unitary if there exists a unitary matrix $U$ such that the action of the evolution operator $\psi$ can be expressed as

$$\psi(P) = U^* PU .$$

For example, quantum walks on graphs and evolutions of wave functions give rise to unitary Markov sources. It is easily seen that a unitary Markov source is nonnegative:

$$\langle P, \psi(U) \rangle = \text{tr}(PU^* PU) = \langle UP, UP \rangle \geq 0 .$$

4.2 Generalized Markov processes and linear representations

We cannot establish an ergodic result for generalized Markov processes $(X_t)$ with finite alphabet $\Sigma$ that is as strong as the statement of Theorem 1. Yet, also a generalized Markov process is asymptotically stationary (in a stochastic rather than quantum state sense).

The key for establishing asymptotic stationarity of $(X_t)$ is to consider the rank $r_k$ of the associated (infinite) matrix of conditional probabilities

$$\mathcal{P} = \begin{bmatrix} p(v|w)_{v,w \in \Sigma^*} \end{bmatrix} .$$

Recall that $p^t(w)$ denotes the probability for the word $w \in \Sigma^*$ to be observed immediately after time $t$. Then one can show
Lemma 4 ([10]). If \( \text{rk} \mathcal{P} < \infty \), then the limit of the expected sample averages

\[
\overline{p}(w) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} p^k(w)
\]

exist for every \( w \in \Sigma^* \). Moreover, the \( \overline{p}(w) \) are the probabilities of a stationary random process \( (X_t) \) with the same alphabet \( \Sigma \).

*Proof.* Let \( \Pi_w \in S \) be the self-adjoint \((N \times N)\)-matrix associated with the word \( w \in \Sigma^* \) and recall the defining property \( p(v|w) = \text{tr}(\sigma^v \Pi_w) \) for all \( v \in \Sigma^* \). The scalar-valued operator \( Q \to s^v(Q) = \text{tr}^v(Q) \) acts linearly on the (real) space \( \mathcal{V} \subseteq S \) relative to which the operators \( \sigma^v \) are defined. Thinking of \( Q = A + iB \in \mathcal{V} \) as the pair \((A,B)\) or real \((N \times N)\)-matrices, \( \mathcal{V} \) can be interpreted as a subspace of the coordinate space \( \mathbb{R}^{4N^2} \). Consequently, we extend \( s^v \) to a scalar-valued linear function on \( \mathbb{R}^{4N^2} \) and represent it by a \((4N^2)\)-dimensional coefficient vector \( s_v \). Writing the \( \Pi_w \)'s as \((4N^2)\)-dimensional column vectors \( \Pi'_w \) and collecting them into the matrix \( \mathcal{P}' = [\Pi'_w] \), we now observe

\[
p(v|w) = s^v(\Pi_w) = s^T_v \Pi'_w \quad (w \in \Sigma^*)
\]

and see that a row of \( \mathcal{P} = [p(v|w)] \) is always a linear combination of the \( 4N^2 \) rows of \( \mathcal{P}' \) (according to the coefficient vector \( s_v \), if the row is labeled \( v \)). Consequently, we have

\[
\text{rk} \mathcal{P} \leq \text{rk} \mathcal{P}' \leq 4N^2 < \infty.
\]

*Proof.* Let \( \Pi_w \in S \) be the self-adjoint \((N \times N)\)-matrix associated with the word \( w \in \Sigma^* \) and recall the defining property

\[
p(v|w) = \text{tr}(\sigma^v \Pi_w) \quad \text{for all} \quad v \in \Sigma^*.
\]

The scalar-valued operator \( Q \to s^v(Q) = \text{tr}^v(Q) \) acts linearly on the (real) space \( \mathcal{V} \subseteq S \) relative to which the operators \( \sigma^v \) are defined. Thinking of \( Q = A + iB \in \mathcal{V} \) as the pair \((A,B)\) or real \((N \times N)\)-matrices, \( \mathcal{V} \) can be interpreted as a subspace of the coordinate space \( \mathbb{R}^{4N^2} \). Consequently, we extend \( s^v \) to a scalar-valued linear function on \( \mathbb{R}^{4N^2} \) and represent it by a \((4N^2)\)-dimensional coefficient vector \( s_v \). Writing the \( \Pi_w \)'s as \((4N^2)\)-dimensional column vectors \( \Pi'_w \) and collecting them into the matrix \( \mathcal{P}' = [\Pi'_w] \), we now observe

\[
p(v|w) = s^v(\Pi_w) = s^T_v \Pi'_w \quad (w \in \Sigma^*)
\]

and see that a row of \( \mathcal{P} = [p(v|w)] \) is always a linear combination of the \( 4N^2 \) rows of \( \mathcal{P}' \) (according to the coefficient vector \( s_v \), if the row is labeled \( v \)). Consequently, we have

\[
\text{rk} \mathcal{P} \leq \text{rk} \mathcal{P}' \leq 4N^2 < \infty.
\]

Theorem 2 shows that the class of generalized Markov processes is equivalent to the class of stochastic process that are known as finitary linearly dependent processes (cf. Gilbert [12] and Ito et al. [15], for example) or concrete observable operator models (cf. Jaeger [16]).

Consider a generalized Markov process \( (X_t) \) with matrix \( \mathcal{P} = [p(v|w)] \) of conditional probabilities. Denote the column with index \( w \) of \( \mathcal{P} \) by \( g_w \) and associate the self-adjoint matrix \( \Pi_w \) with the column vector \( g_w \). This way, the operators \( \sigma^a \) can be viewed as operators on the column space \( \text{lin}\mathcal{P} \) of \( \mathcal{P} \) with respect to the field \( \mathbb{R} \) of scalars. It is straightforward to check that the operators \( \sigma^a \) (and hence the evolution operator \( \psi \)) extend indeed to linear operators

\[
\sigma^a : \text{lin}\mathcal{P} \to \text{lin}\mathcal{P} \quad \text{with the property} \quad \sigma^a g_w = p(a|w)g_{wa}.
\]
In this framework, Lemma 4 says that the following (pointwise) limit exists:

$$G = \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} \psi^k(g)$$  \hspace{1cm} (3)$$

Assuming \( \dim \text{lin} \mathcal{P} = \text{rk} \mathcal{P} = n \), let \( B = \{b_1, \ldots, b_n\} \) be a column basis of \( \mathcal{P} \). Let furthermore \( \pi_w \in \mathbb{R}^n \) denote the coordinate vector of \( g_w \) relative to \( B \) and \( S^a \) the \((n \times n)\)-matrix representing \( \sigma^a \). So we have for all \( w \in \Sigma^* \) and \( a \in \Sigma \),

$$g_w = \sum_{j=1}^{n} \pi_w(j) b_j \quad \text{and} \quad S^a \pi_w = p(a|w) \pi_{wa}.$$  

Notice that the coordinate sums of the \( \pi_w \) are always \( 1^T \pi_w = 1 \) because

$$1 = \sum_{a \in \Sigma} p(a|w) = \sum_{a \in \Sigma} g_w(a) = \sum_{j=1}^{n} \pi_w(j) \sum_{a \in \Sigma} b_j(a) = \sum_{j=1}^{n} \pi_w(j).$$

Similarly, the evolution operator \( \psi \) is represented by the matrix \( M = \sum_{a \in \Sigma} S^a \) with column sums 1.

So we have found a representation of the generalized Markov process \( (X_t) \) in terms of \( n \)-dimensional coordinate vectors and linear operators, where the trace-property becomes a column sum property:

$$\text{tr} S^a \Pi_w = p(a|w) = 1^T S^a \pi_w.$$  

Hence, if \( \pi^0 \) is the coordinate vector of the initial ground state \( g_{\square} \), we have

$$p(w) = 1^T S^w \pi^0 \quad \text{for all} \ w \in \Sigma^*.$$  

5 Markov sources with definite states

The linear representation suggests to interpret a generalized Markov process as being produced by a generalized Markov source with hidden states. We choose

$$\Omega = \{(a,j) \mid a \in \Sigma, 1 \leq j \leq n\}$$

as a representative set for \( N = dm \) hidden states (where \( d = |\Sigma| \)) and consider the information function

$$X : \Omega \to \Sigma \quad \text{such that} \quad X(a,j) = a \quad \text{for all} \ (a,j) \in \Omega.$$  

With each \( S^a = [s_{ij}^a] \in \mathbb{R}^{n \times n} \), we associate the \((N \times N)\)-matrix

$$\overline{S}^a = [s_{(b,i),(c,j)}^a] \quad \text{with} \quad s_{(b,i),(c,j)}^a = \begin{cases} s_{ij}^a & \text{if} \ b = a \\ 0 & \text{if} \ b \neq a \end{cases}$$
and with the initial state representative $\pi^0 \in \mathbb{R}^n$ the $N$-dimensional coordinate vector

$$\pi^0 \text{ with } \pi(a,j) = \pi^0(j)/d.$$ 

By construction, we have now for all $w \in \Sigma^*$,

$$1^T S^w \pi^0 = 1^T S_w \pi^0 = p(w).$$

The Markov process $(X_t)$ can thus be interpreted as emanating from a source with hidden (but definite) states $\Omega$ and the information function $X$. The Markov (hidden state) transition matrix

$$\bar{M} = \sum_{a \in \Sigma} S^a \in \mathbb{R}^{N \times N}$$

has column sums 1 and generally preserves the coefficient sums of coordinate vectors. So we have arrived at a model for Markov sources, which differs from the classical model only in that we allow transition "probabilities" to be possibly negative real numbers.

In particular, all quantum Markov sources can be modeled as generalized Markov sources with possibly negative transition parameters. The source emits the value $X(\omega)$ of the information function $X : \Omega \to \Sigma$ when it is in the hidden state $\omega \in \Omega$.

5.1 Regular sources

Extending the terminology of classical Markov sources we call a source with transition matrix $M$ regular if $M$ admits an orthonormal basis $\{u_1, \ldots, u_N\}$ of eigenvectors $u_i$ with eigenvalues $\lambda_i$ such that $\lambda_1 = 1$ and $|\lambda_i| \neq 1$ holds for all $i \neq 1$. We now write

$$\pi^0 = \sum_{i=1}^N \alpha_i u_i \text{ and } \pi^t = M^t \pi^0 = \sum_{i=1}^N \alpha_i \lambda_i^t u_i = \alpha_1 u_1 + \sum_{i=2}^N \alpha_i \lambda_i^t u_i$$

and recall from Lemma 2(b) that $\|M^t u_i\| = |\lambda_i|^t \|u_i\| = |\lambda_i|^t$ is bounded. Therefore, $\lambda_i \to 0$ holds for all $i \neq 1$ and we conclude $\lim_{t \to \infty} \pi^t = \alpha_1 u_1$. This limit $\alpha_1 u_1$ is an eigenvector of $M$ with eigenvalue $\lambda_1 = 1$ and has component sum 1. Hence it must be the limit state vector:

$$\pi = \alpha_1 u_1 = \lim_{t \to \infty} \pi^t.$$ 

The same argument shows that $\pi$ is independent of the initial state vector $\pi^0$. So we see that the matrix powers $M^t$ converge to a matrix having all columns equal to $\pi$:

$$[\pi, \ldots, \pi] = \lim_{t \to \infty} M^t = M^\infty.$$ 

5.2 Mixing rates

Still assuming a regular Markov source with transition matrix $M$ and set $\Omega$ of hidden states, we turn to the question how fast the ground state vectors $\pi^t = M^t \pi^0$ approach
the limit state $\pi$. Our model now allows an analysis in the spirit of Sinclair and Jer- rum [23], who actually measure the difference between the matrices $M^t$ and $M^\infty$.

With the notation $M^t = [m_{ij}^{(t)}]$, we let $\delta_{ij}(t) = |m_{ij}^{(t)} - \pi_i|$. Choosing the $j$th unit vector $e_j$ as initial state, i.e., $\pi^0 = e_j$, we obtain the $j$th column of $M^t$ as

$$M^t \pi^0 = \pi^t = \pi + \sum_{s=2}^{N} \alpha_s \lambda_s^t u_s .$$

and therefore, in view of the triangle inequality and $|\alpha_s u_s(i)| \leq \|\pi^0\| = 1$,

$$\delta_{ij}(t) = |\sum_{s=0}^{N} \lambda_s \alpha_s u_s(i)| \leq \max_{s \geq 2} |\lambda_s|^t .$$

6 Observable Markov sources

We now think of a Markov source with a set $\Omega = \{1, \ldots, N\}$ of $N$ hidden states as being specified by a (generalized) state transition matrix $M = [m_{ij}]$ with real (but possibly negative) coefficients $m_{ij}$ and the column sum property

$$1 = \sum_{i=1}^{N} m_{ij} \quad (j = 1, \ldots, n) .$$

Given a coordinate vector $w^0 \in \mathbb{R}^N$ with component sum $1^T w^0 = 1$ and a set $\Sigma$, we call the function $X : \Omega \rightarrow \Sigma$ observable (over time) if

$$0 \leq 1^T S^v w^0 \leq 1 \quad \text{for all } v \in \Sigma^* ,$$

where the matrix $S^v$ is defined exactly as in the classical case. So the information function $X$ is said to be observable if its observation results in a (generalized) Markov process $(X_t)$ with classical probabilities $p(w) = 1^T S^w w^0$.

Recall from Section [4] that the expected values of sampling averages exist in the limit if $X$ is observable. If $\Sigma \subseteq \mathbb{R}$, the expectation of the empirical average value,

$$\overline{\mu}(X) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} \mu(X_k) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} \sum_{x \in \Sigma} x \Pr\{X_k = x\}$$

is well-defined in the limit, for example. If one samples for a particular $a \in \Sigma$, the limit yields the expected sampling frequency (and hence the probability in the sense of von Mises [20]):

$$\overline{\mu}(a) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} \Pr\{X_k = a\} .$$
6.1 Compatible observations

This notion of observability can be extended to several information functions. The $\ell$ functions $X_h : \Omega \rightarrow \Sigma_h$ ($h = 1, \ldots, \ell$) are said to be (jointly) compatible with respect to observations over time if the function

$$X : \Omega \rightarrow \prod_{h=1}^\ell \Sigma_h \quad \text{with} \quad X(\omega) = [X_1(\omega), \ldots, X_\ell(\omega)]$$

is observable. It is straightforward to check that arbitrary subsets of compatible functions are compatible. In particular, the individual functions $X_h : \Omega \rightarrow \Sigma_h$ are observable.

In the case of a classical Markov source, the transition matrix $M$ and the initial state vector $w^0 = p^0$ are nonnegative. Hence any collection of functions on $\Omega$ forms a jointly compatible set.

Let $X, Y : \Omega \rightarrow \Sigma \subseteq \mathbb{R}$ be two compatible functions. Then also their product $Z = XY$ is observable. Sampling $X$ and $Y$ over time yields an empirical product with asymptotic expected value

$$\overline{\mathbb{E}}(XY) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^t \sum_{x, y \in \Sigma} xyp_k(x, y).$$

where $p_t(x, y) = \Pr\{X_t = y, Y_t = y\}$. So the expected empirical covariance exists in the limit:

$$\overline{\text{cov}}(X, Y) = \overline{\mathbb{E}}(XY) - \overline{\mathbb{E}}(X)\overline{\mathbb{E}}(Y).$$

If $w^0 = [w_1^0, \ldots, w_N^0]^T$ is an eigenvector of $M$, i.e., $w^0 = Mw^0$, the Markov source is stationary and $p_t(x, y) = p_t(x, y)$ holds for all $t$. In this case, one has

$$\overline{\mathbb{E}}(X) = \sum_{x \in \Sigma} xwp_0(x) = \sum_{i=1}^N X(i)w_i^0 = \mu(X_t) \quad \text{for all } t = 1, 2, \ldots.$$

Similarly, we find $\overline{\mathbb{E}}(Y) = \mu(Y_t)$. The expected value of the empirical covariance, therefore, is just the covariance:

$$\overline{\text{cov}}(X, Y) = \sum_{x, y \in \Sigma} xyp_t(x, y) - \mu(X_t)\mu(Y_t) = \text{cov}(X_t, Y_t) \quad \text{for all } t = 1, 2, \ldots.$$

6.2 Bell’s inequality

The well-known inequality of Bell [4] takes the following form [4] in our context:

**Lemma 5.** Let $X, Y, Z : \Omega \rightarrow \{-1, +1\}$ be arbitrary jointly compatible observations on the stationary Markov source $M$. Then the following inequality holds

$$|\overline{\text{cov}}(X, Y) - \overline{\text{cov}}(Y, Z)| \leq 1 - \overline{\text{cov}}(X, Z). \quad (4)$$
Proof. Any choice of \( x, y, z \in \{-1, +1\} \) satisfies the inequality \( |xy - yz| \leq 1 - xz \). By our compatibility assumption, all \( p_t(x, y, z) = Pr\{X_t = x, Y_t = y, Z_t = z\} \) are nonnegative. So we conclude
\[
|\text{cov}(X_t, Y_t) - \text{cov}(Y_t, Z_t)| = \left| \sum_{x,y,z} (xy - yz)p_t(x, y, z) \right| \leq \sum_{x,y,z} |xy - yz|p_t(x, y, z) \\
\leq \sum_{x,y,z} (1 - xz)p_t(x, y, z) = 1 - \text{cov}(X_t, Z_t).
\]

Note that Bell’s inequality may be violated by a stationary Markov source when \( X, Y, Z \) are pairwise but not jointly compatible. Consider, for example, the trivial (and nonnegative) Markov transition matrix \( M = I \), which always yields a stationary process, and the set of hidden states \( \Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\} \) and functions \( X, Y, Z \) as in the following table:

<table>
<thead>
<tr>
<th>( \omega )</th>
<th>( \omega_1 )</th>
<th>( \omega_2 )</th>
<th>( \omega_3 )</th>
<th>( \omega_4 )</th>
<th>( \omega_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X )</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>( Y )</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>( Z )</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

One can check that \( X, Y, Z \) are pairwise compatible with respect to the vector state
\[
w^0 = [-1/3, 1/3, 1/3, 1/3, 1/3]^T.
\]

The covariances relative to \( w^0 \), however, violate (4):
\[
\text{cov}(X, Y) = +1, \, \text{cov}(Y, Z) = -1/3, \, \text{cov}(X, Z) = +1.
\]

REMARK. There are experimental results that seem to indicate that quantum systems may violate Bell’s inequality (see, e.g., Aspect et al. [2]). This is sometimes interpreted as showing that quantum mechanics does not admit a theory with hidden variables. In our approach, a violation of Bell’s inequality only indicates that the system under consideration is not a stationary classical Markov source. Hidden variables (states) are not excluded if one accepts possibly negative parameters to model general "states" and transitions.

References