

Markov chains with doubly stochastic transition matrices and application to a sequence of non-selective quantum measurements

A. Vourdas

*Department of Computer Science,
University of Bradford,
Bradford BD7 1DP, United Kingdom
a.vourdas@bradford.ac.uk*

A time-dependent finite-state Markov chain that uses doubly stochastic transition matrices, is considered. Entropic quantities that describe the randomness of the probability vectors, and also the randomness of the discrete paths, are studied. Universal convex polytopes are introduced which contain all future probability vectors, and which are based on the Birkhoff-von Neumann expansion for doubly stochastic matrices. They are universal in the sense that they depend only on the present probability vector, and are independent of the doubly stochastic transition matrices that describe time evolution in the future. It is shown that as the discrete time increases these convex polytopes shrink, and the minimum entropy of the probability vectors in them increases. These ideas are applied to a sequence of non-selective measurements (with different projectors in each step) on a quantum system with d -dimensional Hilbert space. The unitary time evolution in the intervals between the measurements, is taken into account. The non-selective measurements destroy stroboscopically the non-diagonal elements in the density matrix. This ‘hermaphrodite’ system is an interesting combination of a classical probabilistic system (immediately after the measurements) and a quantum system (in the intervals between the measurements). Various examples are discussed. In the ergodic example, the system follows asymptotically all discrete paths with the same probability. In the example of rapidly repeated non-selective measurements, we get the well known quantum Zeno effect with ‘frozen discrete paths’ (presented here as a biproduct of our general methodology based on Markov chains with doubly stochastic transition matrices).

I. INTRODUCTION

There are many processes with the Markov property in Science and Engineering and also in Social Sciences, and there is a vast amount of literature that discusses these processes (e.g., [1, 2]). Motivated by an application to quantum systems with non-selective measurements, we study a time-dependent finite-state Markov chain with doubly stochastic transition matrices. Markov chains usually involve the more general row Markov transition matrices.

The novel result in our case with doubly stochastic transition matrices, is the universal convex polytopes. They contain all future probability vectors, and are based on the Birkhoff-von Neumann expansion for doubly stochastic matrices. The universal convex polytopes depend only on the present probability vector, and are independent of the doubly stochastic matrices that describe time evolution in the future. As the discrete time increases these convex polytopes shrink.

In this context, we use entropic quantities to quantify the randomness of the probability vectors. The Kullback-Leibler distance quantifies the distance of the probability vectors from the ‘uncertain’ probability vector (in Eq.(5) below). We also introduce ‘discrete paths’ and quantify their randomness with entropic quantities. Ergodic Markov chains are an important special case, which is studied in this context.

These ideas are applied to a quantum system with d -dimensional Hilbert space (e.g.,[3–5]), on which we perform a sequence of non-selective measurements (with different projectors in each step). Non-selective measurements have been used recently in quantum communications[6, 7], quantum control[8], quantum Zeno effect [9–12], quantum thermodynamics[13, 14], etc.

Between the non-selective measurements our system evolves with unitary transformations. The non-selective measurements destroy stroboscopically the non-diagonal elements in the density matrix, and make the system a probabilistic mixture of states. In each step we use different projectors, and therefore the non-diagonal elements refer to different bases. There is a fundamental difference between a quantum system with superposition of

states, and a classical probabilistic system with a probabilistic mixture of states. The present system belongs to the former category in the intervals between measurements, and to the latter category at the times of the measurements. In this sense it is a ‘hermaphrodite’ system that combines a classical probabilistic system with a quantum system. This is a very different concept from a semiclassical system ($\hbar \rightarrow 0$).

Quantum Mechanics is not a Markovian theory (e.g., [15]). This is related to the non-diagonal terms in the density matrix, which in turn are related to the superposition principle. We note that there are various modifications of Markov models, called ‘quantum Markov models’ (e.g., [16–20]), which are suitable for the description of open quantum systems. Our statement that quantum mechanics is not a Markovian theory refers to the ‘standard’ (non-quantum) Markov models. The difference between classical Markov models and quantum models in the context of probabilities related to social sciences, is discussed in [21].

The quantum system we study is forced by the non-selective measurements to be stroboscopically in a probabilistic mixture of states. Between successive measurements the system evolves in a unitary quantum mechanical manner. We show that the probability vectors associated to the measurements are described by a Markov chain that involves doubly stochastic transition matrices.

In section 2 we introduce briefly doubly stochastic matrices and permutation matrices. In section 3 we discuss time-dependent finite-state Markov chains with doubly stochastic matrices. Entropic quantities are used to describe the randomness of the probability vectors and of the discrete paths (in sections III A, III B, III C). Ergodic Markov chains with doubly stochastic matrices, where the system follows asymptotically all paths with the same probability, are discussed in section III D.

An important aspect of our discussion on Markov chains, is the universal convex polytopes which contain all future probability vectors (in section III E). They are universal in the sense that they depend only on the present probability vector, and they do not depend on the doubly stochastic transition matrices used in the Markov chain in the future. Such polytopes are defined only for the special case of Markov chains with doubly stochastic transition matrices, and for this reason cannot be found in the literature on Markov chains. We show that as the step number increases, the convex polytopes shrink and the minimum entropy of the probability vectors in them increases.

In section 4 we first show that due to the non-diagonal elements in the density matrix, quantum Mechanics is not a Markovian theory (see also [15]). We then consider a sequence of non-selective measurements together with the unitary evolution transformations between them. We show that this can be described as a Markov chain that involves doubly stochastic matrices related to the unitary time evolution matrices. We also define universal convex polytopes that contain all future probability vectors associated with the measurements.

In section 5 we discuss various examples. We consider a system which is initially in a position state, in a momentum state, and in the state $\rho = \frac{1}{d}\mathbf{1}$. We also discuss the homogeneous case, and the ergodic case. We also discuss the special case of rapidly repeated non-selective measurements. This leads to the well known quantum Zeno effect (‘freezing of the paths’), which is presented here with our methodology based on Markov chains with doubly stochastic transition matrices. We conclude in section 6, with a discussion of our results.

II. DOUBLY STOCHASTIC MATRICES

Let \mathfrak{D} be a $d \times d$ matrix with non-negative elements. It is a doubly stochastic matrix if

$$\sum_{a=1}^d \mathfrak{D}(a, b) = \sum_{b=1}^d \mathfrak{D}(a, b) = 1; \quad \mathfrak{D}(a, b) \geq 0. \quad (1)$$

The product of two doubly stochastic matrices is a doubly stochastic matrix. The inverse of a doubly stochastic matrix might not exist, or if it exists it might not be a doubly stochastic matrix. The set of doubly stochastic matrices is a semigroup with respect to matrix multiplication. Physically this means that processes that involve doubly stochastic matrices are irreversible.

The matrix \mathfrak{U} with all elements

$$\mathfrak{U}(a, b) = \frac{1}{d}; \quad \text{rank}(\mathfrak{U}) = 1, \quad (2)$$

is doubly stochastic. For any doubly stochastic matrix \mathfrak{D} we get

$$\mathfrak{U}\mathfrak{D} = \mathfrak{D}\mathfrak{U} = \mathfrak{U}. \quad (3)$$

Let \mathbf{x} be a probability vector

$$\mathbf{x} = (x_0, \dots, x_{d-1}); \quad x_i \geq 0; \quad \sum_{i=0}^{d-1} x_i = 1. \quad (4)$$

The product $\mathbf{x}\mathfrak{D}$ of a probability vector \mathbf{x} (written as a row) times a doubly stochastic matrix \mathfrak{D} (on the right), is also a probability vector (written as a row). We note that the convention in the area of Markov processes is to multiply row vectors with matrices on the right.

A special case of a probability vector, is the most uncertain vector

$$\mathbf{u} = (u_0, \dots, u_{d-1}); \quad u_i = \frac{1}{d}. \quad (5)$$

For any doubly stochastic matrix \mathfrak{D} we get

$$\mathbf{u}\mathfrak{D} = \mathbf{u}. \quad (6)$$

Also for any probability vector \mathbf{x} we get

$$\mathbf{x}\mathfrak{U} = \mathbf{u}. \quad (7)$$

A doubly stochastic matrix has the eigenvalue $e_0 = 1$, and the other eigenvalues e_1, \dots, e_{d-1} have absolute value $|e_i| \leq 1$. The eigenvector corresponding to the eigenvalue $e_0 = 1$, is \mathbf{u}^T . In particular, the matrix \mathfrak{U} has the eigenvalues $e_0 = 1$ and $e_a = 0$ for $a = 1, \dots, d-1$. Therefore any doubly stochastic matrix can be written as

$$\mathfrak{D} = \mathbf{u}^T \mathbf{u} + \sum_{i=1}^{d-1} e_i \mathbf{v}_i^T \mathbf{v}_i = \mathfrak{U} + \sum_{i=1}^{d-1} e_i \mathbf{v}_i^T \mathbf{v}_i; \quad |e_i| \leq 1, \quad (8)$$

where \mathbf{v}_i^T are the eigenvectors corresponding to the eigenvalue e_i .

We define the entropy of a probability vector as

$$E(\mathbf{x}) = - \sum_a x_a \log x_a; \quad 0 \leq E(\mathbf{x}) \leq \log d \quad (9)$$

$E(\mathbf{x}) = 0$ for the most certain vector that has one component equal to one, and the other components equal to zero. $E(\mathbf{x}) = \log d$ for the most uncertain vector \mathbf{u} .

In the set \mathcal{P} of all probability vectors we define the ‘more certain’ preorder[22, 23]:

$$\mathbf{x} \succ \mathbf{y} \leftrightarrow E(\mathbf{x}) \leq E(\mathbf{y}) \quad (10)$$

It is easily seen that

$$\begin{aligned} \mathbf{x} &\succ \mathbf{x} \\ \mathbf{x} &\succ \mathbf{y} \text{ and } \mathbf{y} \succ \mathbf{z} \rightarrow \mathbf{x} \succ \mathbf{z}. \end{aligned} \quad (11)$$

Therefore \succ is a preorder in \mathcal{P} . We note that there are many quantities which are associated with this preorder (Schur concave functions) but in this paper we are interested only in the entropy. It is known [22, 23] that for any doubly stochastic matrix \mathfrak{D} ,

$$\mathbf{x} \succ \mathbf{x}\mathfrak{D}. \quad (12)$$

Also $\mathbf{x} \succ \mathbf{u}$, for any \mathbf{x} .

A. Permutation matrices: a special case of doubly stochastic matrices

Let π be a permutation of the integers $0, \dots, d-1$:

$$(0, 1, \dots, d-1) \xrightarrow{\pi} (\pi(0), \pi(1), \dots, \pi(d-1)). \quad (13)$$

The set of all permutations is the symmetric group $\mathcal{S}[24]$ which has $d!$ elements. Multiplication in this group is the composition:

$$(0, 1, \dots, d-1) \xrightarrow{\pi_1} (\pi_1(0), \pi_1(1), \dots, \pi_1(d-1)) \xrightarrow{\pi_2} (\pi_2[\pi_1(0)], \pi_2[\pi_1(1)], \dots, \pi_2[\pi_1(d-1)]). \quad (14)$$

We will use the notation $(\pi_2 \circ \pi_1)(i) = \pi_2[\pi_1(i)]$. The unity is

$$(0, 1, \dots, d-1) \xrightarrow{1} (0, 1, \dots, d-1). \quad (15)$$

A permutation matrix M_π is a $d \times d$ matrix with elements

$$M_\pi(a, b) = \delta(\pi(a), b); \quad \text{rank}(M_\pi) = d. \quad (16)$$

where δ is the Kronecker delta. Each row and each column have one element equal to 1 and the other $d-1$ elements equal to 0. Permutation matrices are special cases of the doubly stochastic matrices. It is easily seen that

$$M_{\pi_2} M_{\pi_1} = M_{\pi_1 \circ \pi_2}; \quad M_{\mathbf{1}} = \mathbf{1}; \quad M_{\pi^{-1}} = [M_\pi]^{-1} = [M_\pi]^T. \quad (17)$$

The matrices M_π with matrix multiplication form a representation of the symmetric group $\mathcal{S}[24]$.

The Birkhoff-von Neumann theorem states that every doubly stochastic matrix can be expanded in terms of the $d!$ permutation matrices with probabilities $\lambda(\pi)$ as coefficients:

$$\mathfrak{D} = \sum_{\pi} \lambda(\pi) M_\pi; \quad \sum_{\pi} \lambda(\pi) = 1; \quad \lambda(\pi) \geq 0. \quad (18)$$

This expansion is not unique. There is much work on the numerical calculation of the coefficients, which however is not relevant here.

If $\mathfrak{D}_1, \mathfrak{D}_2$ are doubly stochastic matrices, then the $p\mathfrak{D}_1 + (1-p)\mathfrak{D}_2$ where $0 \leq p \leq 1$ is a doubly stochastic matrix. The set of all $d \times d$ doubly stochastic matrices is a convex polytope with the $d!$ permutation matrices (in the symmetric group \mathcal{S}) as vertices. It is known as Birkhoff polytope and is used in areas like Operational Research, Optimization, Enumerative Combinatorics, etc.

General references on convex polytopes are [25, 26], and in a quantum context [27]. Numerical work to find the vertices of a convex polytope is discussed in [28].

III. MARKOV CHAINS WITH DOUBLY STOCHASTIC MATRICES

In a finite-state Markov chain the system is in one of d states which we represent with integers in \mathbb{Z}_d (the integers modulo d). In discrete-time Markov chains we have a sequence of steps (times) $t = 1, 2, \dots$. At each step we have a probability vector (x_0, \dots, x_{d-1}) that gives probabilities for the system to be in the states $0, \dots, d-1$, correspondingly.

A ‘discrete path’ $\wp(1, n)$ of the system between $t = 1$ and $t = n$ is a sequence of n integers in \mathbb{Z}_d :

$$\wp(1, n) = (a_1, \dots, a_n) \in (\mathbb{Z}_d)^n. \quad (19)$$

It shows that the system is in the state $\Sigma = a_1$ at $t = 1$, it is in the state $\Sigma = a_2$ at $t = 2$, etc. There are d^n paths between $t = 1$ and $t = n$.

We assume that at time $t = 1$ the probability vector is

$$\mathbf{x}^{(1)} = (x_0^{(1)}, \dots, x_{d-1}^{(1)}), \quad (20)$$

and that at time $t = 2$ it becomes

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} \mathfrak{D}_{12}. \quad (21)$$

In the literature on Markov chains the transition matrix \mathfrak{D}_{12} , is assumed to be a row Markov matrix. We are interested in the special case where \mathfrak{D}_{12} is a doubly stochastic matrix.

$\mathfrak{D}_{12}(a, b)$ is the probability that the state $\Sigma = a$ at $t = 1$, will become the state $\Sigma = b$ at $t = 2$. Since the state $\Sigma = a$ at $t = 1$ will become some state at $t = 2$, it follows that $\sum_b \mathfrak{D}_{12}(a, b) = 1$, i.e., \mathfrak{D}_{12} is a row Markov matrix. In a classical system there is no reason for the extra requirement $\sum_a \mathfrak{D}_{12}(a, b) = 1$, which makes the matrix doubly stochastic. We will see later that our system is quantum, and in the time interval between $t = 1$ and $t = 2$ evolves with a unitary transformation. In this case we have quantum mechanical conservation of probabilities, which leads to the extra condition $\sum_a \mathfrak{D}_{12}(a, b) = 1$.

Since at $t = 1$ the system is in the state $\Sigma = a$ with probability $x_a^{(1)}$, it follows that the probability that the system will follow the path $\wp(1, 2) = (a, b)$, is

$$q(a, b) = x_a^{(1)} \mathfrak{D}_{12}(a, b); \quad \sum_b q(a, b) = x_a^{(1)}; \quad \sum_a q(a, b) = x_b^{(2)}; \quad \sum_{a,b} q(a, b) = 1. \quad (22)$$

We next assume that at time $t = 3$ the probability vector becomes

$$\mathbf{x}^{(3)} = \mathbf{x}^{(2)} \mathfrak{D}_{23} = \mathbf{x}^{(1)} \mathfrak{D}_{13}; \quad \mathfrak{D}_{13} = \mathfrak{D}_{12} \mathfrak{D}_{23}. \quad (23)$$

Here \mathfrak{D}_{23} is a doubly stochastic transition matrix, and therefore \mathfrak{D}_{13} is a doubly stochastic matrix (as a product of two doubly stochastic matrices). For a time-dependent (non-homogeneous) Markov chain, \mathfrak{D}_{23} is in general different from \mathfrak{D}_{12} . In this case the probability that the system will follow the path $\wp(1, 3) = (a, b, c)$ is

$$q(a, b, c) = x_a^{(1)} \mathfrak{D}_{12}(a, b) \mathfrak{D}_{23}(b, c); \quad \sum_c q(a, b, c) = q(a, b); \quad \sum_{a,b} q(a, b, c) = x_c^{(3)}. \quad (24)$$

We continue in a similar way and at time $t = s$ the probability vector becomes

$$\mathbf{x}^{(s)} = \mathbf{x}^{(s-1)} \mathfrak{D}_{s-1,s} = \mathbf{x}^{(1)} \mathfrak{D}_{1s}; \quad \mathfrak{D}_{1s} = \mathfrak{D}_{12} \dots \mathfrak{D}_{s-1,s}. \quad (25)$$

Here \mathfrak{D}_{1s} is a doubly stochastic transition matrix. In this case the probability that the system will follow the path $\wp(1, s) = (a_1, \dots, a_s)$ is

$$q(a_1, a_2, \dots, a_s) = x_{a_1}^{(1)} \mathfrak{D}_{12}(a_1, a_2) \mathfrak{D}_{23}(a_2, a_3) \dots \mathfrak{D}_{s-1,s}(a_{s-1}, a_s);$$

$$\sum_{a_1, \dots, a_s} q(a_1, a_2, \dots, a_s) = 1. \quad (26)$$

It is easily seen that

$$\sum_{a_2, \dots, a_{s-1}} \mathfrak{D}_{12}(a_1, a_2) \mathfrak{D}_{23}(a_2, a_3) \dots \mathfrak{D}_{s-1,s}(a_{s-1}, a_s) = \mathfrak{D}_{1s}(a_1, a_s). \quad (27)$$

Using it we prove that the path probabilities obey the relations

$$\sum_{a_2, \dots, a_s} q(a_1, a_2, \dots, a_s) = x_{a_1}^{(1)}; \quad \sum_{a_1, \dots, a_{s-1}} q(a_1, a_2, \dots, a_s) = x_{a_s}^{(s)}. \quad (28)$$

They also obey the relations

$$\sum_{a_1, \dots, a_k} q(a_1, \dots, a_k, a_{k+1}, \dots, a_s) = q(a_{k+1}, \dots, a_s); \quad k < s. \quad (29)$$

The probability $q(A)$ that the path $\varphi(1, s) \in A$, where A is a subset of $(\mathbb{Z}_d)^s$ is

$$q(A) = \sum_{(a_1, \dots, a_s) \in A} q(a_1, a_2, \dots, a_s); \quad A \subseteq (\mathbb{Z}_d)^s \quad (30)$$

We use the notation

$$\mathfrak{D}_{k\ell} = \mathfrak{D}_{k,k+1} \mathfrak{D}_{k+1,k+2} \dots \mathfrak{D}_{\ell-1,\ell}; \quad k < \ell. \quad (31)$$

Then

$$\mathfrak{D}_{k\ell} = \mathfrak{D}_{kr} \mathfrak{D}_{r\ell}; \quad k < r < \ell \quad (32)$$

Also

$$\mathbf{x}^{(s)} = \mathbf{x}^{(k)} \mathfrak{D}_{ks}; \quad k < s \quad (33)$$

Knowledge of $\mathbf{x}^{(k)}$ alone determines all the future $\mathbf{x}^{(\ell)}$ with $k < \ell$ (knowledge of the past $\mathbf{x}^{(r)}$ with $r < k$ is not required for the calculation). This ‘lack of memory’ is the Markov property. Given the ‘present’, the ‘future’ is independent of the ‘past’. The present alone determines the future, and the history can be disregarded.

A. Entropy of the probability vectors

In this section we introduce entropic quantities associated with the probability vectors in the Markov chain. Using Eq.(12) we get:

$$\mathbf{u} \prec \dots \prec \mathbf{x}^{(s+1)} \prec \mathbf{x}^{(s)} \prec \dots \prec \mathbf{x}^{(1)}, \quad (34)$$

where \mathbf{u} is the uncertain vector in Eq.(5), or equivalently

$$\log d \geq \dots \geq E[\mathbf{x}^{(s+1)}] \geq E[\mathbf{x}^{(s)}] \geq \dots \geq E[\mathbf{x}^{(1)}] \geq 0. \quad (35)$$

The increase in the entropy is related to the fact that doubly stochastic matrices form a semigroup with respect to multiplication, and therefore the above process is in general irreversible. In the very special case of a Markov chain with permutation matrices (which form the symmetric group), the process is reversible.

The

$$\Delta E(s, s+k) = E[\mathbf{x}^{(s+k)}] - E[\mathbf{x}^{(s)}] \geq 0 \quad (36)$$

is the entropy production when the probability vector $\mathbf{x}^{(s)}$ becomes $\mathbf{x}^{(s+k)}$. The following additivity relation holds:

$$\Delta E(s, s+\ell) = \Delta E(s, s+k) + \Delta E(s+k, s+\ell); \quad k \leq \ell. \quad (37)$$

B. Kullback-Leibler distance

We quantify how close is the probability vector $\mathbf{x}^{(s)}$ to the uncertain vector \mathbf{u} , using the relative entropy

$$S(\mathbf{x}^{(s)}||\mathbf{u}) = \sum x_i^{(s)} \log \frac{x_i^{(s)}}{u_i} = \log d - E(\mathbf{x}^{(s)}); \quad 0 \leq S(\mathbf{x}^{(s)}||\mathbf{u}) \leq \log d. \quad (38)$$

It is known as Kullback-Leibler distance between the two probability vectors [29, 30] (although mathematically it does not have the properties of a distance, for example $S(\mathbf{x}^{(s)}||\mathbf{u}) \neq S(\mathbf{u}||\mathbf{x}^{(s)})$). From Eq.(35) follows that as the step s increases the $S(\mathbf{x}^{(s)}||\mathbf{u})$ decreases, which shows that $\mathbf{x}^{(s)}$ gets closer to the uncertain vector \mathbf{u} :

$$0 \leq \dots \leq S(\mathbf{x}^{(s+1)}||\mathbf{u}) \leq S(\mathbf{x}^{(s)}||\mathbf{u}) \leq \dots \leq S(\mathbf{x}^{(1)}||\mathbf{u}) \leq \log d. \quad (39)$$

If for some s we get $\mathbf{x}^{(s)} = \mathbf{u}$ then subsequently, for all $t > s$ we get $\mathbf{x}^{(t)} = \mathbf{u}$.

C. Entropy of the discrete paths

The probability that the system will follow the path $\varphi(1, s) = (a_1, \dots, a_s)$ is $q(a_1, a_2, \dots, a_s)$ in Eq.(45). We define the following entropy related to these probabilities:

$$E_\varphi(1, s) = -\frac{1}{s} \sum_{a_1, \dots, a_s} q(a_1, a_2, \dots, a_s) \log q(a_1, a_2, \dots, a_s). \quad (40)$$

The number of paths is d^s and we use the normalization factor $\frac{1}{s}$ so that the maximum value of $E_\varphi(1, s)$ is $\log d$:

$$0 \leq E_\varphi(1, s) \leq \log d. \quad (41)$$

$E_\varphi(1, s)$ quantifies the randomness in the path of the system. Similar quantities are used in the context of dynamical systems [31, 32], and here we use them in the first instance for Markov chains and later for quantum systems. It is easily seen that

$$E_\varphi(1, s) = \frac{1}{s} \left\{ E[\mathbf{x}^{(1)}] + \sum_{a_1} x_{a_1}^{(1)} \mathcal{E}(a_1) \right\}$$

$$\mathcal{E}(a_1) = - \sum_{a_2, \dots, a_s} [\mathfrak{D}_{12}(a_1, a_2) \dots \mathfrak{D}_{s-1, s}(a_{s-1}, a_s)] \log [\mathfrak{D}_{12}(a_1, a_2) \dots \mathfrak{D}_{s-1, s}(a_{s-1}, a_s)] \quad (42)$$

D. Ergodic Markov chain with doubly stochastic matrices

There are various definitions of ergodic Markov chains in the literature, for the general case that the transition matrices are row Markov matrices [33–37]. Some of these definitions are weaker than others. The general idea is that in the large n limit, all rows in the transition matrices $\mathfrak{D}_{k, k+n}(a, b)$ should be ‘almost equal’ to a ‘stationary probability vector’ (which does not depend on the row number a).

In our Markov chains the transition matrices are doubly stochastic, and it follows that the stationary probability vector can only be the most uncertain probability vector \mathbf{u} in Eq.(5). \mathbf{u} is a stationary probability vector for any doubly stochastic matrix (Eq.(6)). In this case the general idea is that in the large n limit, the transition matrices $\mathfrak{D}_{1, n}(a, b)$ should be ‘almost equal’ to \mathfrak{U} (in Eq.(2)).

A Markov chain with doubly stochastic matrices is ergodic if

$$\lim_{n \rightarrow \infty} \mathfrak{D}_{1, n}(a, b) = \frac{1}{d} \quad (43)$$

In this case for every infinitesimal $\epsilon > 0$, there exist n_0 such that for all $n > n_0$

$$\frac{1}{d} - \epsilon \leq \mathfrak{D}_{1,n}(a, b) \leq \frac{1}{d} + \epsilon. \quad (44)$$

We note that:

- for every pair of states (a, b) there exists a non-zero probability that the state $\Sigma = a$ at the step $t = 1$, will be transformed to the state $\Sigma = b$ at some step $t = n$. This is called irreducibility.
- The probability $\mathfrak{D}_{1,n}(a, a)$ is non-zero for all states $\Sigma = a$ and for all $n > n_0$. This is called aperiodicity.

From Eq.(43) follows that the probability for the path $\wp(1, n) = (a_1, \dots, a_{n_0}, \dots, a_n)$ is

$$\begin{aligned} A \left(\frac{1}{d} - \epsilon \right)^{n-n_0} &\leq q(a_1, \dots, a_{n_0}, \dots, a_n) \leq A \left(\frac{1}{d} + \epsilon \right)^{n-n_0} \\ A &= x_{a_1}^{(1)} \mathfrak{D}_{12}(a_1, a_2) \dots \mathfrak{D}_{n_0-1, n_0}(a_{n_0-1}, a_{n_0}) \end{aligned} \quad (45)$$

and we find that

$$\lim_{n \rightarrow \infty} \left[q(a_1, \dots, a_n) - \frac{1}{d^n} \right] = 0. \quad (46)$$

This means that in the ergodic Markov chain and in the large n limit, the system follows all paths $\wp(1, n)$ with the same probability $\frac{1}{d^n}$. Consequently the entropy of the paths takes the largest value $\log d$:

$$\lim_{n \rightarrow \infty} E_{\wp}(1, n) = \log d. \quad (47)$$

Example III.1. We consider a homogeneous Markov chain with

$$\mathfrak{D}_{12} = \mathfrak{D}_{23} = \dots = \mathfrak{D}_{s-1, s} = \mathfrak{U}. \quad (48)$$

According to our definition this is an ergodic Markov chain. We consider two initial probability vectors.

- The first one is $\mathbf{x}^{(1)} = \mathbf{u}$. Then

$$q(a_1, a_2, \dots, a_s) = \frac{1}{d^s}; \quad E_{\wp}(1, s) = \log d. \quad (49)$$

The system follows all paths with the same probability $\frac{1}{d^s}$.

- The second initial probability vector is $\mathbf{x}^{(1)} = (1, 0, \dots, 0)$. Then

$$q(a_1, a_2, \dots, a_s) = \frac{\delta(a_1, 0)}{d^{s-1}}; \quad E_{\wp}(1, s) = \frac{s-1}{s} \log d. \quad (50)$$

In the large s limit the system follows all paths with the same probability:

$$q(a_1, a_2, \dots, a_s) \rightarrow \frac{1}{d^s} \quad E_{\wp}(1, s) \rightarrow \log d. \quad (51)$$

Example III.2. We consider a homogeneous Markov chain case with

$$\mathfrak{D}_{12} = \mathfrak{D}_{23} = \dots = \mathfrak{D}. \quad (52)$$

Therefore

$$\mathbf{x}^{(s+1)} = \mathbf{x}^{(1)} \mathfrak{D}^s. \quad (53)$$

We next assume that all eigenvalues of \mathfrak{D} apart from the eigenvalue $e_0 = 1$, have absolute value $|e_i| < 1$. Then as $s \rightarrow \infty$ the \mathfrak{D}^s has the eigenvalues $e_0^s = 1$, and e_i^s with $|e_i^s| \rightarrow 0$. Using Eq.(8), we get

$$\lim_{s \rightarrow \infty} \mathfrak{D}^s(a, b) = \frac{1}{d}. \quad (54)$$

Therefore this is ergodic Markov chain. The $e_{\max} = \max\{|e_1|, \dots, |e_{d-1}|\}$ defines how quickly we will reach the limit. The $(e_{\max})^s$ should be close to zero, and this leads to

$$s \gg \frac{1}{-\log e_{\max}}. \quad (55)$$

E. Universal convex polytopes

In this section we define convex polytopes related to the probability vectors. They are for Markov chains with doubly stochastic matrices, and for this reason they have not been studied in the literature that uses general row Markov matrices.

Below we sometimes represent a probability vector with its first $d - 1$ independent components (the last component is not independent). This is indicated with a ‘hat’ in the notation:

$$\mathbf{x} = (x_0, \dots, x_{d-1}) \rightarrow \hat{\mathbf{x}} = (x_0, \dots, x_{d-2}); \quad \sum_{i=0}^{d-2} x_i \leq 1. \quad (56)$$

For example, the probability vector $\mathbf{x} = (0, \dots, 0, 1)$ is represented with $\hat{\mathbf{x}} = (0, \dots, 0)$.

We define the simplex Δ_{d-1}

$$\Delta_{d-1} = \left\{ (t_0, \dots, t_{d-2}) \mid \sum_{i=0}^{d-2} t_i \leq 1 \right\}, \quad (57)$$

which contains all probability vectors $\hat{\mathbf{x}}$. It is a $(d - 1)$ -dimensional convex polytope, with the d vertices

$$\text{vert}(\Delta_{d-1}) = \{(0, \dots, 0), (0, \dots, 0, 1), \dots, (1, 0, \dots, 0)\}. \quad (58)$$

We consider a Markov chain with the probability vector $\mathbf{x}^{(1)}$ (in Eq.(20)) at time $t = 1$. We define the

$$\mathbf{x}_{\pi}^{(1)} = \mathbf{x}^{(1)} M_{\pi} = \left(x_{\pi^{-1}(0)}^{(1)}, \dots, x_{\pi^{-1}(d-1)}^{(1)} \right); \quad E(\mathbf{x}_{\pi}^{(1)}) = E(\mathbf{x}^{(1)}). \quad (59)$$

The components of $\mathbf{x}_{\pi}^{(1)}$ are a permutation of the components of $\mathbf{x}^{(1)}$. There are $d!$ such vectors, but some of them might be equal to each other.

We define the convex polytope $\mathcal{A}[\mathbf{x}^{(1)}]$ that contains all the probability vectors that are convex combinations of the probability vectors $\hat{\mathbf{x}}_{\pi}^{(1)}$ (represented as the hat notation indicates with their first $d - 1$ components):

$$\mathcal{A}[\mathbf{x}^{(1)}] = \left\{ \sum_{\pi} \lambda(\pi) \hat{\mathbf{x}}_{\pi}^{(1)}; \quad \lambda(\pi) \geq 0; \quad \sum_{\pi} \lambda(\pi) = 1 \right\}. \quad (60)$$

We note that:

- $\mathcal{A}[\mathbf{x}^{(1)}]$ is a universal convex polytope in the sense that it depends only on the probability vector $\mathbf{x}^{(1)}$. It does not depend on the doubly stochastic matrices D_t with $t > 1$ that will be used in this Markov chain in the future. In other words Markov chains with different doubly stochastic matrices, which start with the same probability vector $\mathbf{x}^{(1)}$, have the same convex polytope $\mathcal{A}[\mathbf{x}^{(1)}]$.

- $\mathcal{A}[\mathbf{x}^{(1)}]$ is a subset of Δ_{d-1} . Also $\mathcal{A}[\mathbf{x}^{(1)}]$ contains the vector $\hat{\mathbf{u}} = (\frac{1}{d}, \dots, \frac{1}{d})$. We easily prove this by taking in Eq.(60) only the d permutations which are ‘shifts’

$$\left(x_0^{(1)}, \dots, x_{d-2}^{(1)}\right), \left(x_1^{(1)}, \dots, x_{d-1}^{(1)}\right), \left(x_2^{(1)}, \dots, x_{d-1}^{(1)}, x_0^{(1)}\right), \dots, \left(x_{d-1}^{(1)}, x_0^{(1)}, \dots, x_{d-3}^{(1)}\right) \quad (61)$$

with coefficients $1/d$. The vector $\hat{\mathbf{u}}$ is only one point in the polytope and can be viewed as the zero-dimensional simplex Δ_0 . Therefore

$$\Delta_0 \subseteq \mathcal{A}[\mathbf{x}^{(1)}] \subseteq \Delta_{d-1}. \quad (62)$$

- The minimum entropy of the probability vectors in $\mathcal{A}[\mathbf{x}^{(1)}]$ is $E(\mathbf{x}^{(1)})$. Indeed, the concavity property of the entropy shows that every probability vector $\mathbf{y} \in \mathcal{A}[\mathbf{x}^{(1)}]$ has entropy

$$\log d \geq E(\mathbf{y}) = E\left(\sum_{\pi} \lambda(\pi) \mathbf{x}_{\pi}^{(1)}\right) \geq \sum_{\pi} \lambda(\pi) E(\mathbf{x}_{\pi}^{(1)}) = E(\mathbf{x}^{(1)}). \quad (63)$$

The importance of $\mathcal{A}[\mathbf{x}^{(1)}]$ lies in the fact that all future probability vectors in the Markov chain belong to $\mathcal{A}[\mathbf{x}^{(1)}]$.

Proposition III.3. *In the above Markov chain:*

- (1) Every future probability vector $\mathbf{x}^{(s+1)}$ with $s > 0$, belongs to the polytope $\mathcal{A}[\mathbf{x}^{(1)}]$ in Eq.(60).
- (2) every probability vector $\mathbf{y} \in \mathcal{A}[\mathbf{x}^{(1)}]$ is a potential future probability vector, in the sense that there exists a doubly stochastic matrix \mathfrak{D} such that $\mathbf{y} = \mathfrak{D}\mathbf{x}^{(1)}$.

Proof. (1) Using the Birkhoff-von Neumann expansion

$$\mathfrak{D}_{1s} = \sum_{\pi} \alpha_{1s}(\pi) M_{\pi} \quad (64)$$

where $\alpha_{1s}(\pi)$ are probabilities, we express $\mathbf{x}^{(s+1)}$ as

$$\mathbf{x}^{(s+1)} = \mathbf{x}^{(1)} \mathfrak{D}_{1s} = \sum_{\pi} \alpha_{1s}(\pi) \mathbf{x}_{\pi}^{(1)}. \quad (65)$$

Therefore $\mathbf{x}^{(s+1)}$ belongs to the convex polytope $\mathcal{A}[\mathbf{x}^{(1)}]$.

- (2) Let $\mathbf{y} \in \mathcal{A}[\mathbf{x}^{(1)}]$. Then by definition there exist probabilities $\lambda(\pi)$ such that $\mathbf{y} = \sum_{\pi} \lambda(\pi) \mathbf{x}_{\pi}^{(1)}$. But then

$$\mathbf{y} = \mathbf{x}^{(1)} \mathfrak{D}; \quad \mathfrak{D} = \sum_{\pi} \lambda(\pi) M_{\pi}. \quad (66)$$

□

In analogy to $\mathcal{A}[\mathbf{x}^{(1)}]$ we define a convex polytope for every s :

$$\Delta_0 \subseteq \mathcal{A}[\mathbf{x}^{(s)}] = \left\{ \sum_{\pi} \lambda(\pi) \hat{\mathbf{x}}_{\pi}^{(s)}; \lambda(\pi) \geq 0; \sum_{\pi} \lambda(\pi) = 1 \right\} \subseteq \Delta_{d-1}. \quad (67)$$

$\mathcal{A}[\mathbf{x}^{(s)}]$ contains all future probability vectors $\mathbf{x}^{(t)}$ with $t > s$, and therefore

$$\Delta_0 \subseteq \dots \subseteq \mathcal{A}[\mathbf{x}^{(s+1)}] \subseteq \mathcal{A}[\mathbf{x}^{(s)}] \subseteq \dots \subseteq \mathcal{A}[\mathbf{x}^{(1)}] \subseteq \Delta_{d-1}. \quad (68)$$

As the step number s increases the convex polytope shrinks, and its dimension might stay the same or it might decrease. Also the minimum entropy of the probability vectors in the polytope increases, as s increases. (Eq.(34)).

$\mathcal{A}[\mathbf{x}^{(s)}]$ is a universal convex polytope in the sense that it depends only on the probability vector $\mathbf{x}^{(s)}$. It does not depend on the doubly stochastic matrices D_t with $t > s$ that will be used in this Markov chain in the future. Of course the shrinking of this polytope in the future, depends on the doubly stochastic matrices D_t with $t > s$.

Example III.4. For $d = 3$, we consider a Markov chain with the probability vector

$$\mathbf{x}^{(1)} = (0.2, 0.3, 0.5), \quad (69)$$

at $t = 1$. We have six permutations of the vector $\mathbf{x}^{(1)}$. We write the six vectors $\widehat{\mathbf{x}}_{\pi}^{(1)}$ as the matrix

$$M = \begin{pmatrix} 0.2 & 0.3 \\ 0.2 & 0.5 \\ 0.3 & 0.2 \\ 0.5 & 0.2 \\ 0.3 & 0.5 \\ 0.5 & 0.3 \end{pmatrix}. \quad (70)$$

We use this matrix with the program 'conhull' in MATLAB and we get the outer polytope in Fig.1, which is 2-dimensional and has six vertices. The polytope is universal in the sense that all future probability vectors (their first two components) belong to this polytope, regardless of the doubly stochastic matrices that will be used in this Markov chain. For example, we see from Fig.1 that in the future we will never get the probability vector $(0.4, 0.5, 0.1)$.

For the next step, we consider the doubly stochastic matrix

$$\mathfrak{D}_{12} = \begin{pmatrix} 0.1 & 0.3 & 0.6 \\ 0.4 & 0.2 & 0.4 \\ 0.5 & 0.5 & 0 \end{pmatrix} \quad (71)$$

and we get the probability vector

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)}\mathfrak{D}_{12} = (0.39, 0.37, 0.24), \quad (72)$$

at $t = 2$. We then repeat the above process and we find that the polytope shrinks into the inner polytope in Fig.1.

The Kullback-Leibler distance of the probability vectors $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ from \mathbf{u} is

$$S(\mathbf{x}^{(1)}||\mathbf{u}) = 0.069; \quad S(\mathbf{x}^{(2)}||\mathbf{u}) = 0.021. \quad (73)$$

Here we used natural logarithms and therefore the result is in nats.

Example III.5. For $d = 4$, we consider a Markov chain, with probability vector

$$\mathbf{x}^{(1)} = (0.1, 0.2, 0.3, 0.4), \quad (74)$$

at $t = 1$. We have 24 'permutation vectors' $\widehat{\mathbf{x}}_{\pi}^{(1)}$. We use the same process as in the previous example and we get the 3-dimensional convex polytope in Fig.2.

Example III.6. In a Markov chain at $t = 1$ the probability vector is $\mathbf{x}^{(1)} = (1, 0, \dots, 0)$. There are only d permutations of this vector which are different from each other, and which are convexly independent vectors (none of them is a convex combination of the others). Then $\mathcal{A}[\mathbf{x}^{(1)}]$ is the $(d - 1)$ -simplex:

$$\mathcal{A}[\mathbf{x}^{(1)}] = \Delta_{d-1} \quad (75)$$

We note that this contains all probability vectors, so in this particular example the convex polytope does not give any useful information.

Example III.7. If $\mathbf{x}^{(1)} = \mathbf{u}$ (defined in Eq.(5)) all permutations $\widehat{\mathbf{x}}_{\pi}^{(1)}$ of this vector are equal to each other. The convex polytope in this case is just one point:

$$\mathcal{A}[\mathbf{x}^{(1)}] = \Delta_0. \quad (76)$$

IV. A SEQUENCE OF NON-SELECTIVE QUANTUM MEASUREMENTS

A. Quantum Mechanics is not a Markovian theory

We consider a quantum system (qudit) with variables in \mathbb{Z}_d (the integers modulo d). H_d is the d -dimensional Hilbert space describing this system. $|f\rangle$ where $f \in \mathbb{Z}_d$, is an orthonormal basis in H_d which we call basis of position states. For later use, we also consider the Fourier transform

$$\mathcal{F} = \frac{1}{\sqrt{d}} \sum_{f,g} \exp\left(\frac{2\pi fg}{d}\right) |f\rangle\langle g| \quad (77)$$

Acting with \mathcal{F} on the basis of position states, we get a dual basis that we call momentum states:

$$|f\rangle_F = \mathcal{F}|f\rangle \quad (78)$$

We assume that initially the system is described by a density matrix ρ with elements $\rho(f, g) = \langle f|\rho|g\rangle$ (in the basis of position states). The density matrix evolves in time with a unitary matrix V into $\sigma = V^\dagger \rho V$ which has matrix elements

$$\sigma(f, g) = \sum_h [V(h, f)]^* \rho(h, k) V(k, g) \quad (79)$$

If the density matrix ρ is diagonal ($\rho(f, g) = 0$ for $f \neq g$) then the diagonal elements of σ are given by

$$\sigma(f, f) = \sum_h \rho(h, h) \mathfrak{D}(h, f); \quad \mathfrak{D}(h, f) = |V(h, f)|^2. \quad (80)$$

Since V is unitary matrix, it follows that $\mathfrak{D}(h, f)$ is a doubly stochastic matrix and then Eq.(80) is analogous to Eq.(21) in Markov chains. But σ also has the non-diagonal elements

$$\sigma(f, g) = \sum_h [V(h, f)]^* \rho(h, h) V(h, g); \quad f \neq g. \quad (81)$$

It is seen that if the initial density matrix ρ is diagonal, time evolution gives a density matrix σ which has diagonal elements given by the Markov relation in Eq.(80), but which also has non-diagonal elements $\sigma(f, g)$. Non-diagonal elements in the density matrix are intimately related to the superposition principle, and are at the heart of Quantum Mechanics. It is therefore clear that Quantum Mechanics is not a Markovian theory (see also [15]).

Quantum evolution is a reversible process associated with a group of unitary transformations, while a Markov chain is an irreversible process associated with a semigroup. A generalisation of Markov theory called quantum Markov theory, suitable for the description of quantum systems and their interaction with the environment, has been studied in [16–20].

Quantum systems can be described with Markov chain if we have a mechanism of destroying the non-diagonal elements and introduce irreversibility. We will do this stroboscopically with a sequence of non-selective measurements. Our formalism below is a Markovian stroboscopic description of the system, immediately after the non-selective measurements. In the time intervals between the measurements the system acquires off-diagonal elements in the density matrix, and cannot be described in a Markovian way.

B. Non-selective measurements on qudits

We consider the quantum system (qudit) described in section IV A. Let \mathfrak{P}_f be the orthogonal projectors of rank one:

$$\mathfrak{P}_f = |f\rangle\langle f|; \quad \sum_{f=0}^{d-1} \mathfrak{P}_f = \mathbf{1}. \quad (82)$$

Measurements with \mathfrak{P}_f at $t = 1$ on a system described with the density matrix ρ , will give ‘yes’ with probability

$$x_f^{(1)} = \text{Tr}[\rho\mathfrak{P}_f] = \rho(f, f); \quad \sum_{f=0}^{d-1} x_f^{(1)} = 1. \quad (83)$$

The probability vector

$$\mathbf{x}^{(1)} = (x_0^{(1)}, \dots, x_{d-1}^{(1)}) \quad (84)$$

describes the outcome from the measurements \mathfrak{P}_f (with $f = 0, \dots, d-1$).

In the case of a non-selective measurement (where we do not look at the outcome of the measurement) the system will be described after the measurement, with the density matrix

$$\rho_1 = \sum_f \mathfrak{P}_f \rho \mathfrak{P}_f. \quad (85)$$

The non-selective measurement destroys the off-diagonal terms $\mathfrak{P}_f \rho \mathfrak{P}_g$ with $f \neq g$. For orthogonal projectors of rank one (like the \mathfrak{P}_f), this can be also be written as

$$\rho_1 = \sum_f x_f^{(1)} \mathfrak{P}_f = \sum_f \rho(f, f) |f\rangle\langle f|. \quad (86)$$

In the basis of position states, ρ_1 is a diagonal matrix which has the probability vector in Eq.(84) in the diagonal:

$$\rho_1 = \text{diag}(x_0^{(1)}, \dots, x_{d-1}^{(1)}). \quad (87)$$

The von Neumann entropy of a density matrix ρ is $E(\rho) = -\text{Tr}(\rho \log \rho)$. Since ρ_1 is a diagonal matrix, we get

$$E(\rho_1) = E(\mathbf{x}^{(1)}). \quad (88)$$

Diagonal density matrices can be represented by the probability vectors in their diagonals, and in analogy to $\mathcal{A}[\mathbf{x}^{(1)}]$ we define the convex polytope $\mathcal{B}(\rho_1)$ of these diagonal density matrices. $\mathcal{A}[\mathbf{x}^{(1)}]$ has been defined in terms of probability vectors, and $\mathcal{B}(\rho_1)$ in terms of diagonal density matrices that have these probability vectors in their diagonals. Then

$$\Delta_0 \subseteq \mathcal{B}(\rho_1) \subseteq \Delta_{d-1}. \quad (89)$$

Here Δ_{d-1} contains all diagonal $d \times d$ density matrices, and Δ_0 is a zero-dimensional simplex (one point) related to the density matrix $\rho = \frac{1}{d} \mathbf{1}$.

C. A sequence of non-selective measurements as a Markov chain

We next assume that this system evolves between $t = 1$ and $t = 2$ with the unitary matrix V_{12} (which has matrix elements $V_{12}(f, g)$ in the basis of position states). At $t = 2$ we get the density matrix

$$\tilde{\rho}_2 = \sum_f x_f^{(1)} \mathfrak{P}_f^{(2)}; \quad \mathfrak{P}_f^{(2)} = V_{12}^\dagger \mathfrak{P}_f V_{12}. \quad (90)$$

On the density matrix $\tilde{\rho}_2$, we perform again a non-selective measurement with different projectors which we write as $W_2 \mathfrak{P}_f W_2^\dagger$, where W_2 is a unitary matrix. We then get

$$\begin{aligned} \rho_2 &= \sum_f \text{Tr}(\tilde{\rho}_2 W_2 \mathfrak{P}_f W_2^\dagger) W_2 \mathfrak{P}_f W_2^\dagger = \sum_{f,g} x_g^{(1)} \mathfrak{D}_{12}(g, f) W_2 \mathfrak{P}_f W_2^\dagger; \quad g, f = 0, \dots, d-1 \\ \mathfrak{D}_{12}(g, f) &= \text{Tr}(\mathfrak{P}_g^{(2)} W_2 \mathfrak{P}_f W_2^\dagger) = |(V_{12} W_2)(g, f)|^2. \end{aligned} \quad (91)$$

Since $V_{12}W_2$ is a unitary matrix, $\mathfrak{D}_{12}(g, f)$ is a doubly stochastic transition matrix. The fact that our system evolves quantum mechanically between successive measurements with a unitary transformation (which in turn is related to conservation of probabilities), leads to doubly stochastic transition matrices. A system which is classical between $t = 1$ and $t = 2$, will be described with general row Markov transition matrices.

We rewrite Eq.(91) as

$$\rho_2 = \sum_f x_f^{(2)} W_2 \mathfrak{P}_f W_2^\dagger; \quad x_f^{(2)} = \sum_g x_g^{(1)} \mathfrak{D}_{12}(g, f) = \sum_g \rho(g, g) \mathfrak{D}_{12}(g, f). \quad (92)$$

The $\mathbf{x}^{(2)} = \mathbf{x}^{(1)} \mathfrak{D}_{12}$ is also a probability vector related to the second measurement. ρ_2 in the basis $W_2|f\rangle$ is the diagonal matrix:

$$\rho_2 = \text{diag}(x_0^{(2)}, \dots, x_{d-1}^{(2)}). \quad (93)$$

We next assume that this system evolves between $t = 2$ and $t = 3$ with unitary transformations V_{23} and we get the density matrix

$$\tilde{\rho}_3 = \sum_f x_f^{(2)} \mathfrak{P}_f^{(3)}; \quad \mathfrak{P}_f^{(3)} = V_{23}^\dagger W_2 \mathfrak{P}_f W_2^\dagger V_{23}. \quad (94)$$

On the density matrix $\tilde{\rho}_3$, we perform again a non-selective measurement with different projectors $W_3 \mathfrak{P}_f W_3^\dagger$ where W_3 is a unitary matrix, and we get

$$\begin{aligned} \rho_3 &= \sum_f \text{Tr}(\tilde{\rho}_3 W_3 \mathfrak{P}_f W_3^\dagger) W_3 \mathfrak{P}_f W_3^\dagger = \sum_{f,g} x_g^{(2)} \mathfrak{D}_{23}(g, f) W_3 \mathfrak{P}_f W_3^\dagger = \sum_f x_f^{(3)} W_3 \mathfrak{P}_f W_3^\dagger \\ \mathfrak{D}_{23}(g, f) &= \text{Tr}(\mathfrak{P}_g^{(3)} W_3 \mathfrak{P}_f W_3^\dagger) = |(W_2^\dagger V_{23} W_3)(g, f)|^2 \\ \mathbf{x}^{(3)} &= \mathbf{x}^{(2)} \mathfrak{D}_{23} \end{aligned} \quad (95)$$

$\mathfrak{D}_{23}(g, f)$ is a doubly stochastic matrix. We rewrite Eq.(95) as

$$\begin{aligned} \rho_3 &= \sum_{f,g,h} x_h^{(1)} \mathfrak{D}_{12}(h, g) \mathfrak{D}_{23}(g, f) W_3 \mathfrak{P}_f W_3^\dagger = \sum_{f,h} x_h^{(1)} \mathfrak{D}_{13}(h, f) W_3 \mathfrak{P}_f W_3^\dagger \\ \mathfrak{D}_{13}(h, f) &= \sum_g \mathfrak{D}_{12}(h, g) \mathfrak{D}_{23}(g, f) = \sum_g |(V_{12} W_2)(h, g)|^2 |(W_2^\dagger V_{23} W_3)(g, f)|^2 \end{aligned} \quad (96)$$

Here $\mathfrak{D}_{13} = \mathfrak{D}_{12} \mathfrak{D}_{23}$ is the product of two doubly stochastic matrices, and therefore it is a doubly stochastic matrix. ρ_3 is the diagonal matrix:

$$\rho_3 = \text{diag}(x_0^{(3)}, \dots, x_{d-1}^{(3)}). \quad (97)$$

We repeat this process many times. It is a sequence of non-selective measurements with unitary transformations that describe time evolution, between them. The unitary transformations are in general different in each step. The projectors related to the measurements are also in general different in each step. On the density matrix $\tilde{\rho}_s$, we perform a non-selective measurement with the projectors $W_s \mathfrak{P}_f W_s^\dagger$ where W_s is a unitary matrix. Schematically we describe this process as follows:

$$\begin{aligned} \rho &\xrightarrow{\text{meas}} \rho_1 \xrightarrow{V_{12}} \tilde{\rho}_2 = V_{12}^\dagger \rho_1 V_{12} \xrightarrow{\text{meas}} \rho_2 \xrightarrow{V_{23}} \tilde{\rho}_3 = V_{23}^\dagger \rho_2 V_{23} \xrightarrow{\text{meas}} \rho_3 \xrightarrow{V_{34}} \dots \\ V_{s,s+1}^\dagger &\tilde{\rho}_{s+1} = V_{s,s+1}^\dagger \rho_{s,s+1} V_{s,s+1} \xrightarrow{\text{meas}} \rho_{s+1} \longrightarrow \dots \end{aligned} \quad (98)$$

Here

$$\begin{aligned} \rho_{s+1} &= \sum_{f,g} x_f^{(s+1)} W_{s+1} \mathfrak{P}_f W_{s+1}^\dagger = \sum_{f,g} x_g^{(s)} \mathfrak{D}_{s,s+1}(g, f) W_{s+1} \mathfrak{P}_f W_{s+1}^\dagger \\ \mathfrak{D}_{s,s+1}(g, f) &= |(W_s^\dagger V_{s,s+1} W_{s+1})(g, f)|^2 \end{aligned} \quad (99)$$

It is seen that

$$\begin{aligned}\mathbf{x}^{(s+1)} &= \mathbf{x}^{(s)} \mathfrak{D}_{s,s+1} = \mathbf{x}^{(k)} \mathfrak{D}_{k,s+1}; \quad k < s + 1 \\ \rho_{s+1} &= \text{diag}(x_0^{(s+1)}, \dots, x_{d-1}^{(s+1)}).\end{aligned}\tag{100}$$

where we use the notation of Eq.(31).

Entropic inequalities in Eqs(34),(35) hold here also. We get the ‘entropic staircase’:

$$\log d \geq \dots \geq E(\rho_{s+1}) \geq E(\tilde{\rho}_{s+1}) = E(\rho_s) \geq \dots \geq E(\tilde{\rho}_2) = E(\rho_1) \geq E(\rho).\tag{101}$$

Unitary time evolution is a reversible process that keeps the entropy constant. The increases in the entropy are related to the measurements, and make the process irreversible.

We also define probabilities for discrete paths (as in Eq.(45)) by looking stroboscopically at the system at the times after the measurements. The corresponding entropy is given by Eq.(40).

We note that using different projectors in each step, simply changes the evolution matrix (e.g., from $V_{s,s+1}$ to $W_s^\dagger V_{s,s+1} W_{s+1}$ in Eq.(99)). In the rest of the paper we assume that all projectors used in the measurements are the same, i.e., that $W_s = \mathbf{1}$.

D. Universal convex polytopes

In analogy to $\mathcal{B}(\rho_1)$ we define a convex polytope $\mathcal{B}(\rho_s)$ for every s . They are analogous to the polytopes $\mathcal{A}[\mathbf{x}^{(s)}]$, and we get here results analogous to those in section III E:

- $\mathcal{B}(\rho_s)$ is universal in the sense that it depends only on ρ_s , and it does not depend on the future time evolution matrices V_t with $t > s$.
- $\mathcal{B}(\rho_s)$ contains all future diagonal density matrices ρ_t with $t > s$ (after each non-selective measurement).
- The minimum entropy of all density matrices in $\mathcal{B}(\rho_s)$ is $E(\rho_s)$.
- As the discrete time s increases, the convex polytope shrinks:

$$\Delta_0 \subseteq \dots \subseteq \mathcal{B}(\rho_{s+1}) \subseteq \mathcal{B}(\rho_s) \subseteq \dots \subseteq \mathcal{B}(\rho_1) \subseteq \Delta_{d-1},\tag{102}$$

and the minimum entropy of the density matrices in it increases.

E. Destruction of the non-diagonal terms with non-selective measurements and the Markov property

We have explained in section IV A that time evolution in quantum mechanics is non-Markovian because of the non-diagonal terms. The process between two non-selective measurements can be described as a Markov chain, because the non-selective measurements destroy the non-diagonal elements.

In order to see this in a different way we compare and contrast the process

$$\rho \xrightarrow{\text{meas}} \rho_1 \xrightarrow{V_{12}} \tilde{\rho}_2 = V_{12}^\dagger \rho_1 V_{12} \xrightarrow{\text{meas}} \rho_2 \xrightarrow{V_{23}} \tilde{\rho}_3 = V_{23}^\dagger \rho_2 V_{23} \xrightarrow{\text{meas}} \rho_3\tag{103}$$

with the case where the second non-selective measurement does not take place:

$$\rho \xrightarrow{\text{meas}} \rho_1 \xrightarrow{V_{12}} \tilde{\rho}_2 = V_{12}^\dagger \rho_1 V_{12} \longrightarrow \sigma_2 = \tilde{\rho}_2 \xrightarrow{V_{23}} \tilde{\sigma}_3 = V_{23}^\dagger \sigma_2 V_{23} \xrightarrow{\text{meas}} \sigma_3\tag{104}$$

In this case

$$\begin{aligned}\sigma_3 &= \sum_{f,h} x_h^{(1)} \check{\mathfrak{D}}_{12}(h, f) \mathfrak{P}_f \\ \check{\mathfrak{D}}_{12}(h, f) &= |(V_{12}V_{23})(h, f)|^2 = \left| \sum_g V_{12}(h, g)V_{23}(g, f) \right|^2\end{aligned}\quad (105)$$

$\check{\mathfrak{D}}_{12}(h, f)$ is a doubly stochastic matrix. We compare Eqs(96),(105) and we note that

$$\check{\mathfrak{D}}_{12}(h, f) = \mathfrak{D}_{12}(h, f) + \sum_{g \neq g'} V_{12}(h, g)V_{23}(g, f)V_{12}(h, g')V_{23}(g', f) \quad (106)$$

The difference between $\check{\mathfrak{D}}_{12}(h, f)$ and $\mathfrak{D}_{12}(h, f)$ is non-diagonal terms which are absent in the latter case because the non-selective measurement destroys them.

V. SPECIAL CASES

In this section we consider special cases of the general theory in section IV.

A. Initial state is a position state

We assume that the initial state of the system is the position state $|r\rangle$. We perform a non-selective measurement with the projectors \mathfrak{P}_f and we get the probability vector $\mathbf{x}^{(1)}$:

$$x_f^{(1)} = \delta(r, f). \quad (107)$$

After the non-selective measurement the system is described with the density matrix ρ_1 :

$$\rho_1 = \text{diag}(0, \dots, 0, 1, 0, \dots, 0). \quad (108)$$

Here the (r, r) -element of ρ_1 is one. The corresponding entropy is

$$E(\rho_1) = E(\mathbf{x}^{(1)}) = 0. \quad (109)$$

The corresponding convex polytope $\mathcal{B}(\rho_1)$ is the $(d-1)$ -simplex:

$$\mathcal{B}(\rho_1) = \Delta_{d-1}. \quad (110)$$

The system evolves in time with unitary transformations V_{12} and we get the density matrix

$$\tilde{\rho}_2 = V_{12}^\dagger |r\rangle \langle r| V_{12}. \quad (111)$$

On the density matrix $\tilde{\rho}_2$, we perform again a non-selective measurement with the same projectors \mathfrak{P}_f and we get

$$x_f^{(2)} = \mathfrak{D}_{12}(r, f); \quad \mathfrak{D}_{12}(r, f) = |V_{12}(r, f)|^2 \quad (112)$$

After the non-selective measurement the system is described with the diagonal density matrix ρ_2 :

$$\rho_2 = \text{diag}(|V_{12}(r, 0)|^2, \dots, |V_{12}(r, d-1)|^2). \quad (113)$$

We continue in the same way and we get

$$\begin{aligned}\mathbf{x}^{(s+1)}(g_{s+1}) &= \mathfrak{D}_{1,s+1}(r, g_{s+1}) \\ \mathfrak{D}_{1,s+1}(r, g_{s+1}) &= \sum_{g_2, \dots, g_s} |V_{12}(r, g_2)|^2 |V_{23}(g_2, g_3)|^2 \dots |V_{s,s+1}(g_s, g_{s+1})|^2.\end{aligned}\quad (114)$$

After the non-selective measurement the system is described with the density matrix

$$\rho_{s+1} = \text{diag}(\mathfrak{D}_{1,s+1}(r, 0), \dots, \mathfrak{D}_{1,s+1}(r, d-1)). \quad (115)$$

The probability that the system will follow the discrete path $\wp(1, s+1) = (g_1, g_2, \dots, g_{s+1})$ is

$$q(g_1, g_2, \dots, g_s) = \delta(r, g_1) |V_{12}(g_1, g_2)|^2 |V_{23}(g_2, g_3)|^2 \dots |V_{s,s+1}(g_s, g_{s+1})|^2. \quad (116)$$

B. Initial state is a momentum state

We assume that the initial state of the system is the momentum state $|r\rangle_F$. We perform a non-selective measurement with the projectors \mathfrak{P}_f and we get the probability vector:

$$\mathbf{x}^{(1)} = \mathbf{u} = \left(\frac{1}{d}, \dots, \frac{1}{d}\right) \quad (117)$$

After the non-selective measurement the system is described with the density matrix ρ_1 :

$$\rho_1 = \text{diag}\left(\frac{1}{d}, \dots, \frac{1}{d}\right). \quad (118)$$

The corresponding convex polytope $\mathcal{B}(\rho_1)$ is the 0-simplex (one point):

$$\mathcal{B}(\rho_1) = \Delta_0. \quad (119)$$

Then for any unitary evolution matrices V_{12}, V_{23}, \dots , we get

$$\mathbf{x}^{(s+1)} = \mathbf{u}; \quad \rho_{s+1} = \text{diag}\left(\frac{1}{d}, \dots, \frac{1}{d}\right); \quad \mathcal{B}(\rho_s) = \Delta_0, \quad (120)$$

for all $s = 0, 1, 2, \dots$. Also

$$E(\rho_{s+1}) = E(\mathbf{x}^{(s+1)}) = \log d. \quad (121)$$

The probability that the system will follow the discrete path $\wp(1, s+1) = (g_1, g_2, \dots, g_{s+1})$ depends on the unitary evolution matrices V_{12}, V_{23}, \dots , and it is given by

$$q(g_1, g_2, \dots, g_s) = \frac{1}{d} |V_{12}(g_1, g_2)|^2 |V_{23}(g_2, g_3)|^2 \dots |V_{s,s+1}(g_s, g_{s+1})|^2. \quad (122)$$

In this example all the probability vectors $\mathbf{x}^{(s+1)}$ are the same, but the probabilities of the various paths are different from each other.

C. Initial state is $\rho = \frac{1}{d}\mathbf{1}$

We assume that the initial density matrix is

$$\rho = \frac{1}{d}\mathbf{1}. \quad (123)$$

In this case we get results analogous to those in the previous example.

D. Ergodic case

We consider a quantum system with $d = 3$ and Hamiltonian

$$h = \begin{pmatrix} 1 & 2 & -i \\ 2 & 2 & i \\ i & -i & 1 \end{pmatrix} \quad (124)$$

We perform non-selective measurements as described earlier, at times $t = 1, 2, \dots$. In this case the time evolution matrices between successive measurements are

$$V_{12} = V_{23} = \dots = \exp(iht); \quad t = 1. \quad (125)$$

From this follows that

$$\mathfrak{D}_{12} = \mathfrak{D}_{23} = \dots = \mathfrak{D} = \begin{pmatrix} 0.232 & 0.223 & 0.545 \\ 0.223 & 0.551 & 0.226 \\ 0.545 & 0.226 & 0.229 \end{pmatrix}. \quad (126)$$

The eigenvalues of this matrix are $e_0 = 1$, $e_1 = 0.325$, $e_2 = -0.315$.

If the probability vector at $t = 1$ is $\mathbf{x}^{(1)}$, then at $t = s$ the probability vector is

$$\mathbf{x}^{(s+1)} = \mathbf{x}^{(1)} \mathfrak{D}^s. \quad (127)$$

As $s \rightarrow \infty$ the \mathfrak{D}^s goes close to \mathfrak{U} , and the $\mathbf{x}^{(s+1)}$ close to \mathbf{u} . In this example, for $s > 5$ the $\mathfrak{D}^s \approx \mathfrak{U}$ and the $\mathbf{x}^{(s+1)} \approx \mathbf{u}$. Therefore the system is ergodic, and for large s it follows all discrete paths with the same probability $1/3^s$.

E. Rapidly repeated non-selective measurements: quantum Zeno effect and freezing of the paths

We consider the homogeneous case with rapidly repeated non-selective measurements. Then \mathfrak{D} is infinitesimally close to $\mathbf{1}$, and it can be written as $\mathfrak{D} = \mathbf{1} + \mathfrak{E}$ where

$$\mathfrak{E} = \begin{pmatrix} -\epsilon(0,0) & \epsilon(0,1) & \dots & \epsilon(0,d-1) \\ \epsilon(1,0) & -\epsilon(1,1) & \dots & \epsilon(1,d-1) \\ \vdots & \vdots & \vdots & \vdots \\ \epsilon(d-1,0) & \epsilon(d-1,1) & \dots & -\epsilon(d-1,d-1) \end{pmatrix} \quad (128)$$

$$\sum_{b \neq a} \epsilon(a,b) - \epsilon(a,a) = \sum_{a \neq b} \epsilon(a,b) - \epsilon(b,b) = 0. \quad (129)$$

Here $\epsilon(a,b)$ are infinitesimal non-negative numbers. For example, we assume that the time evolution matrix for one of the intervals, is

$$V = \exp(iht) \approx \mathbf{1} + iht \quad (130)$$

where h is the Hamiltonian (with matrix elements $h(a,b)$) and t is an infinitesimal time interval. Then

$$\epsilon(a,a) = 1 - [h(a,a)t]^2; \quad \epsilon(a,b) = |h(a,b)t|^2; \quad a \neq b. \quad (131)$$

Then $\mathfrak{D}^s \approx \mathbf{1} + s\mathfrak{E}$ and

$$x_a^{(s+1)} = x_a^{(1)} [1 - s\epsilon(a,a)] + s \sum_{b \neq a} x_b^{(1)} \epsilon(b,a) \quad (132)$$

If ϵ_{\max} is the maximum of all $\epsilon(a, b)$, then for large s which obeys the inequality

$$1 \ll s \ll \frac{1}{\epsilon_{\max}} \quad (133)$$

the probability vector $\mathbf{x}^{(s+1)}$ is almost equal to $\mathbf{x}^{(1)}$:

$$\mathbf{x}^{(s+1)} \approx \mathbf{x}^{(1)}. \quad (134)$$

Consequently the corresponding universal convex polytope remains the same.

We also calculated the probabilities for discrete paths given in Eq.(45). We find that the probability that $\wp(1, s+1)$ is the ‘frozen path’ (a, \dots, a) is

$$q(a, \dots, a) = x_a^{(1)} [\mathfrak{D}(a, a)]^s = x_a^{(1)} [1 - s\epsilon(a, a)] \approx x_a^{(1)} \quad (135)$$

The probability for any path that involves at least two different states is almost zero, because the $\mathfrak{D}(a, b)$ with $a \neq b$ is infinitesimal. For example, for the path where the first k states are a , and the rest of the states are b , the probability is

$$\begin{aligned} q(a, \dots, a, b, \dots, b) &= x_a^{(1)} [\mathfrak{D}(a, a)]^{k-1} \mathfrak{D}(a, b) [\mathfrak{D}(b, b)]^{s-k} \\ &= x_a^{(1)} [1 - (k-1)\epsilon(a, a)] \epsilon(a, b) [1 - (s-k)\epsilon(b, b)] \\ &\approx x_a^{(1)} \epsilon(a, b) \approx 0 \end{aligned} \quad (136)$$

If there are more than two different states in the path, then the corresponding probability involves product of many infinitesimals and is even more close to zero. For large s (which obeys the inequality (133)) the discrete path entropy (in Eq.(40)) goes to zero:

$$E_\wp(1, s+1) = -\frac{1}{s+1} \sum_a q(a, \dots, a) \log q(a, \dots, a) = -\frac{1}{s+1} \sum_a x_a^{(1)} \log x_a^{(1)} \rightarrow 0. \quad (137)$$

It is seen that for time intervals that satisfy the inequality in Eq.(133), rapidly repeated non-selective measurements ‘freeze’ (stroboscopically) the paths of the system. This is the well known quantum Zeno effect [9–12], presented here in the language of Markov chains with doubly stochastic matrices, that has been used in this paper. The fact that the path entropy goes to zero, is also an indication of the ‘freezing’ of the paths.

VI. DISCUSSION

We studied Markov chains with doubly stochastic transition matrices. Entropic quantities have been used to describe the randomness of the probability vectors and of the discrete paths. We have also introduced the universal convex polytopes in Eq.(67) which contain all future probability vectors. We have shown that as the step number (discrete time) increases the convex polytopes shrink (Eq.(68)), and the minimum entropy of the probability vectors in them increases. Ergodic Markov chains have been studied in section III D.

Quantum Mechanics is not a Markovian theory. This is due to the non-diagonal elements in the density matrix, which are related to the superposition principle. We have discussed this in detail in sections IV A and IV E. However non-selective measurements destroy the non-diagonal elements. We have considered a quantum system which is forced stroboscopically by non-selective measurements, to be in a classical probabilistic mixture of states. Between the measurements the system evolves unitarily in a quantum mechanical way. We have shown that this system is described by time-dependent finite-state Markov chains with doubly stochastic transition matrices. We have explained that the quantumness of the system between successive measurements, makes the transition matrices doubly stochastic. We note again here that generalisations of the classical Markov models called quantum Markov models, which are suitable for open quantum systems have been studied in [16–20].

We applied our results for Markov chains with doubly stochastic transition matrices, to our quantum system with the non-selective measurements. In this quantum context, we have also introduced universal convex polytopes related to the diagonal density matrices after the measurements. They contain all future probability vectors related to the measurements. We have shown that as the discrete time increases the convex polytopes shrink (Eq.(102)), and the minimum entropy of the probability vectors in them increases.

Various examples have been discussed in section V. We have considered a system which is initially in a position state, in a momentum state, and in the state $\rho = \frac{1}{d}\mathbf{1}$. In the ergodic case the system follows asymptotically all discrete paths with the same probability. For rapidly repeated non-selective measurements we have shown how our approach which is based on Markov chains with doubly stochastic matrices, leads to the well known quantum Zeno effect.

Nonselective measurements on a quantum system, have recently been used in the literature for various objectives (quantum communications[6, 7], quantum control[8], quantum Zeno effect [9–12], quantum thermodynamics[13, 14], etc). In the present paper a sequence of non-selective measurements has been studied as a ‘classical-quantum system’, which immediately after the measurements is a classical probabilistic system, and in the intervals between successive measurements is a quantum system that evolves with a unitary transformation. We have shown that it is a rich system that exhibits various phenomena in various limits.

-
- [1] J.R. Norris, ‘Markov chains’ (Cambridge Univ. Press, Cambridge, 1997)
 - [2] N. Privault, ‘Understanding Markov chains’ (Springer, Berlin, 2013)
 - [3] A. Vourdas, Rep. Prog. Phys, 67,267 (2004)
 - [4] A. Vourdas, ‘Finite and profinite quantum systems’ (Springer, Berlin, 2017)
 - [5] T. Durt, B.G. Englert, I. Bengtsson, K. Życzkowski, Int. J. Quant. Comp. 8, 535 (2010)
 - [6] A. Kalev, A. Mann, M. Revzen, Phys. Rev. Lett. 110, 260502 (2013)
 - [7] B.A. Nguyen, Adv. Nat. Sci: Nanosci.Nanotechnology 9, 025001 (2018)
 - [8] A. Pechen, N. Ilin, F. Shuang, H. Rabitz, Phys. Rev. A74, 052102 (2006)
 - [9] B. Misra, E.C.G. Sudarshan, J. Math. Phys. 18, 756 (1977)
 - [10] Y. Aharonov, M. Vardi, Phys. Rev. D21, 2235 (1980)
 - [11] W.M. Itano, D.J. Heinzen, J.J. Bollinger, D.J. Wineland, Phys. Rev. A41, 2295 (1990)
 - [12] M. Majid, A.Z. Chaudhry, Sci. Rep. 8, 14887 (2018)
 - [13] P. Kammerlander, J. Anders, Sci. Rep. 6, 22174 (2016)
 - [14] M.H. Mohammady, A. Auffeves, J. Anders, Commun. Phys. 3, 89 (2020)
 - [15] D.T. Gillespie, Phys. Rev. A49, 1607 (1994)
 - [16] L. Acardi, Adv. Math. 20, 329 (1976)
 - [17] C.W. Gardiner, ‘Quantum Noise’ (Springer, Berlin, 1991)
 - [18] H.P. Breuer, F. Petruccione, ‘The theory of open quantum systems’ (Oxford Univ. Press, Oxford, 2002)
 - [19] S. Gudder, J. Math. Phys. 49, 072105 (2008)
 - [20] A. Dhahri, C.K. Ko, Y.J. Yoo, J. Stat. Phys. 176, 1272 (2019)
 - [21] J.R. Busemeyer, P.D. Kvam, T.J. Pleskac, Sci. Rep. 9, 18025 (2019)
 - [22] A.W. Marshal, I. Olkin ‘Inequalities: theory of majorization and its applications’ (Academic Press, New York, 1979)
 - [23] T. Ando, Linear Algebra Appl. 118, 163 (1989)
 - [24] B.E. Sagan ‘The Symmetric group’ (Springer, Berlin, 2001)
 - [25] G. Ziegler, ‘Lectures on polytopes’ (Springer, Berlin, 1995)
 - [26] B. Grunbaum, ‘Convex polytopes’ (Springer, Berlin, 2003)
 - [27] I. Bengtsson, K. Życzkowski, ‘Geometry of quantum states’ (Cambridge Univ. Press, Cambridge, 2006)
 - [28] T.H. Matheiss, D.S. Rubin, Math. Operations Research 5,167 (1980)
 - [29] S. Kullback, R.A. Leibler, Ann. Math. Statistics, 22, 79 (1951)
 - [30] T. Cover, J. Thomas ‘Elements of Information Theory’ (Wiley, New York, 1991)
 - [31] I.P. Cornfeld, S.V. Fomin, Y.G. Sinai, ‘Ergodic systems’ (Springer, Berlin, 1982)
 - [32] P. Gaspard, J. Stat. Phys. 117, 599 (2004)
 - [33] D. Vere-Jones, Quart. J.Math. 13, 7 (1962)
 - [34] J. L.Teugels, J. Applied Prob. 9, 466 (1972)
 - [35] D. Isaacson, R.L. Tweedie, J. Applied Prob. 15, 87 (1978)

- [36] K. Chan *Adv. Applied Prob.* 21, 702 (1989)
- [37] A. Hordijk, F. Spieksma, *Adv. Applied Prob.* 24, 343 (1992)

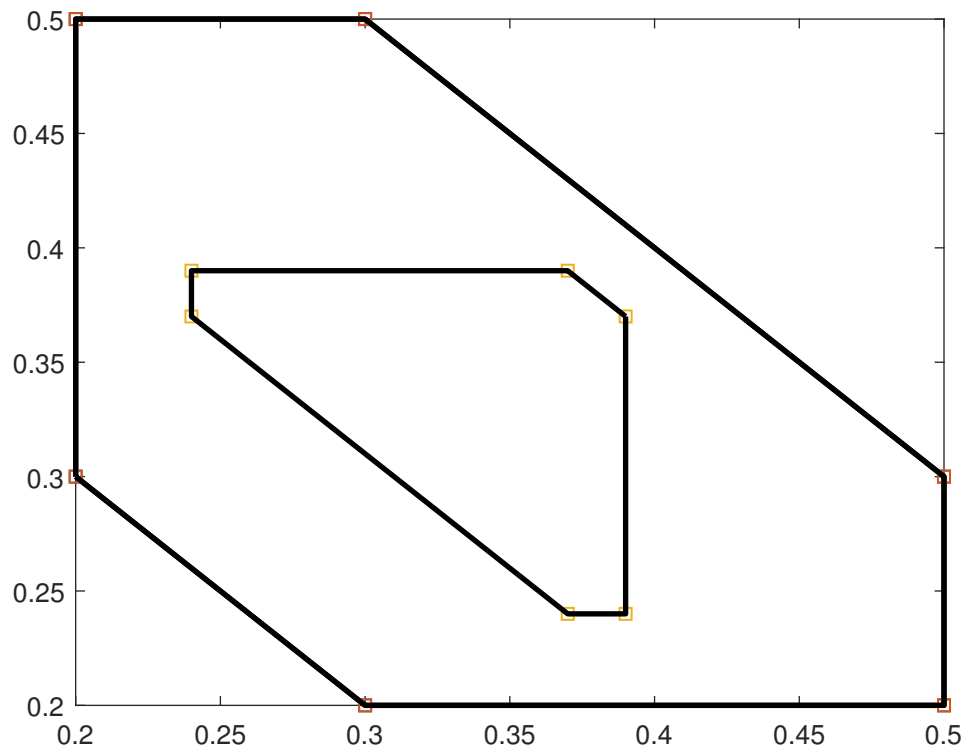


FIG. 1: The outer polytope is the convex polytope for a Markov chain, with the probability vector in Eq.(69). With the doubly stochastic matrix in Eq.(71), we get in the next step the probability vector in Eq.(72), and then the convex polytope shrinks into the inner one.

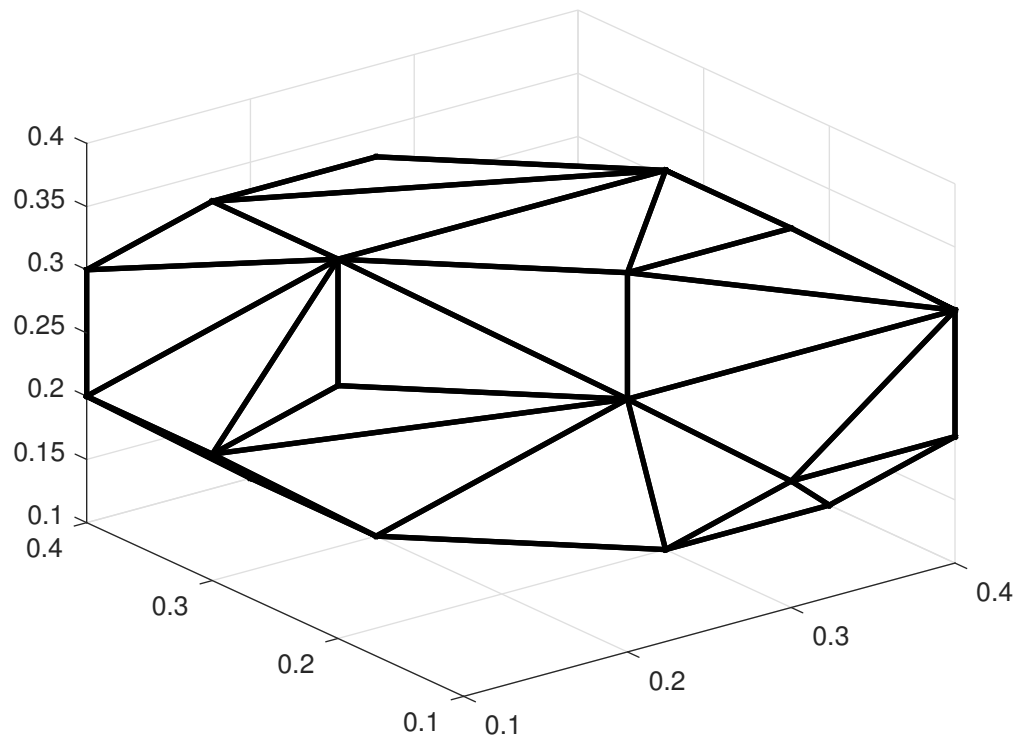


FIG. 2: The 3-dimensional convex polytope for a Markov chain, with the probability vector in Eq.(74).