# 2D Markov Chains

Ettore Fornasini
*Department of Electronics and Computer Science*
*University of Padova*
*6 / A, via Gradenigo*
*35131 Padova, Italy*

Submitted by Richard A. Brualdi

ABSTRACT

The separation property that characterizes the dynamics of Markov chains is extended to a class of discrete 2D models where the time support, given by the discrete plane $\mathbb{Z}\times\mathbb{Z}$, is partially ordered by the product of the orderings. The paper analyzes the matrix representation structure of the probability transition map in a 2D Markov chain and some properties of the associated characteristic polynomial in two variables. These allow one to show how the long-term behavior depends on the intersections between the variety of the characteristic polynomial and the distinguished boundary of the unit closed bidisk.

## 1. INTRODUCTION

Consider any finite homogeneous Markov chain with $n$ states $S_1, S_2, \dots, S_n$. The transitions from one state to another occur at times $\dots, 0, 1, \dots,$ and the probabilistic picture of possible changes at each step is provided by a stochastic matrix $\mathbf{A} \in \mathbb{R}_+^{n\times n}$. Once we know the probabilities

$$\begin{bmatrix} x_1(t) & x_2(t) & \cdots & x_n(t) \end{bmatrix} := \mathbf{x}(t)$$

of the various states at time $t$, the probabilities after one step are the components of the row vector

$$\mathbf{x}(t+1) = \mathbf{x}(t)\mathbf{A}. \tag{1.1}$$




655 Avenue of the Americas, New York, NY 10010 0024-3795/90/$3.50

So, knowing the outcome of the last experiment, we can neglect any other information we have about the past in predicting the future. This separation property, embodied by Equation (1.1), makes it quite natural to look at Markov chains as to a special subclass of positive linear dynamical systems that evolve autonomously on the set

$$X=\left\{\mathbf{x}\in\mathbb{R}_+^n \,\middle|\, \sum_{h=1}^{n} x_h=1\right\} \tag{1.2}$$

of $n$-dimensional probability vectors.

During the last few years a considerable research effort has been devoted to dynamical patterns that evolve in the discrete plane $\mathbb{Z}\times\mathbb{Z}$, partially ordered by the product of the orderings

$$(r,s)\leqslant(h,k) \quad \text{iff} \quad r\leqslant h \text{ and } s\leqslant k. \tag{1.3}$$

The causality constraints that (1.3) naturally induces on the dynamical patterns imply that the configuration attained at $(h,k)$ only depends on configurations and input values at $(r,s)\leqslant(h,k)$.

Autonomous 2D systems [1–3] constitute the easiest nontrivial instance of these dynamical behaviors. Here the local configuration $\mathbf{x}(h+1,k+1)$ is linearly determined by the nearest past configurations $\mathbf{x}(h,k+1)$ and $\mathbf{x}(h+1,k)$. We therefore have the following first-order updating equation:

$$\mathbf{x}(h+1,k+1)=\mathbf{x}(h,k+1)\mathbf{A}^{(1)}+\mathbf{x}(h+1,k)\mathbf{A}^{(2)}, \tag{1.4}$$

where $\mathbf{x}$ is an $n$-dimensional real-valued row vector and $\mathbf{A}^{(1)},\mathbf{A}^{(2)}$ are $n\times n$ real matrices.

In some way, the separation property we have already recognized for Markov chains is inherited by the system (1.4) in a two-dimensional environment. Actually, the computation of the local configuration at $(h+1,k+1)$ doesn't require any information about system history in the "past cone" $\{(r,s)<(h+1,k+1)\}$, with the exception of the nearest points $(h,k+1)$ and $(h+1,k)$.

So, although no particular probability meaning is associated with the local vector $\mathbf{x}$ in the general theory of 2D systems, it seems rather natural to obtain a 2D theory of Markov chains by introducing suitable constraints in Equation (1.4). These must guarantee that any pair of probability vectors $\mathbf{x}(h,k+1)$ and $\mathbf{x}(h+1,k)$ leads in turn to a new probability vector at $(h+1,k+1)$, so that the components of $\mathbf{x}(h+1,k+1)$ can be viewed as probabilities of the various states at point $(h+1,k+1)$.

Quite recently multidimensional Markov models (hidden Markov mesh random fields) have been considered in the image-processing literature, with the purpose of developing coherent approaches to problems of both image segmentation and model acquisition [4]. The problems that will be addressed in this paper are quite different. Our main interest will consist into a system-theoretic description of 2D Markov chains and an outline of some of their internal properties.

The first property is that concerning the algebraic structure of the matrices $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}$ of transition probabilities. It essentially states that the pair $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)})$ can be written as $(a\mathbf{P},(1-a)\mathbf{Q})$, where $\mathbf{P}$ and $\mathbf{Q}$ are stochastic matrices and $0 \leqslant a \leqslant 1$.

The second property is a remarkable restriction on the variety $\mathscr{V}(\Delta)$ of the characteristic 2D polynominal

$$\Delta(z_1,z_2) = \det\left(\mathbf{I} - \mathbf{A}^{(1)}z_1 - \mathbf{A}^{(2)}z_2\right). \tag{1.5}$$

It will be shown that $\mathscr{V}(\Delta)$ intersects the unit closed polydisk $\mathscr{P}_1$ only at some points of its distinguished boundary $\mathscr{T}_1$.

A third result comes under the heading of model analysis and establishes a remarkable connection between the intersection $\mathscr{V}(\Delta) \cap \mathscr{T}_1$ and the long-term behavior of the probability vectors $\mathbf{x}(h,k)$. An interesting question we shall answer in this context is the following: does there exist a probability vector $\mathbf{w}$ such that $\mathbf{x}(h,k)$ approaches $\mathbf{w}$ as $h+k$ tends to infinity? This result provides a significant qualitative conclusion that can be inferred about the behavior of a 2D Markov chain even though the values of the parameters may not be known precisely.

## 2. THE STRUCTURE OF A 2D MARKOV CHAIN

By a 2D Markov chain $\mathscr{M}$ with $n$ states $S_1, S_2, \dots, S_n$ we will mean:

(1) an *autonomous 2D system*

$$\mathbf{x}(h+1,k+1) = \mathbf{x}(h,k+1)\mathbf{A}^{(1)} + \mathbf{x}(h+1,k)\mathbf{A}^{(2)} \tag{2.1}$$

of dimension $n$, with the property that $\mathbf{x}(h+1,k+1)$ is a probability row vector for every pair of probability row vectors $\mathbf{x}(h,h+1)$ and $\mathbf{x}(h+1,k)$;

(2) a sequence of *initial probability vectors*

$$\mathscr{X}_0 = \{\mathbf{x}(h,k) \mid (h,k) \in \mathscr{C}_0,\ \mathbf{x}(h,k) \in X\}, \tag{2.2}$$

where

$$\mathscr{C}_0 = \{(h,k) \in \mathbb{Z}\times\mathbb{Z} \mid h+k=0\} \tag{2.3}$$

is a separation set in $\mathbb{Z}\times\mathbb{Z}$, and $x_i(h,k)$, $i=1,2,\ldots,n$, denotes the probability that $S_i$ is the state of the system at the initial point $(h,k)$.

The pair $(\mathbf{A}^{(1)},\mathbf{A}^{(2)})$ determines the probabilistic behavior of the system, once the probability distributions are given at every point of $\mathscr{C}_0$. Note that the shape of the separation set could have been chosen quite differently from (2.3); however, assuming $\mathscr{C}_0$ to be a diagonal straight line in $\mathbb{Z}\times\mathbb{Z}$ will simplify much notation in the sequel.

A basic question concerning Equation (2.1) is the following: if $\mathbf{x}(h,k+1)$ and $\mathbf{x}(h+1,k)$ are probability vectors, but otherwise arbitrary, under what circumstances can one be certain that the new vector $\mathbf{x}(h+1,k+1)$ will also be of the same type? A first partial result is the following:

LEMMA 2.1. *Let* $\mathbf{P}$ *and* $\mathbf{Q}$ *be* $n\times n$ *stochastic matrices, and consider any real number* $a$ *in the interval* $[0,1]$. *Then* $\mathbf{A}^{(1)}=a\mathbf{P}$, $\mathbf{A}^{(2)}=(1-a)\mathbf{Q}$ *are matrices of a 2D Markov chain.*

*Proof.* We only need to show that $\mathbf{v}a\mathbf{P}+\mathbf{w}(1-a)\mathbf{Q}$ is a probability vector whenever $\mathbf{v}$ and $\mathbf{w}$ are. This is clear, since $\mathbf{vP}$ and $\mathbf{wQ}$ are probability vectors and $X$ is a convex set. ∎

A natural question arises whether the structure considered in the above lemma is in some sense canonical for 2D Markov chains. In order to study this problem, we need a preliminary result, concerning the uniqueness of the representation (2.1). Actually, giving an $n$-state 2D Markov chain $\mathscr{M}$ essentially reduces to assigning a one-step transition-probability map

$$\pi : X\times X \to X \tag{2.4}$$

via the restriction to $X\times X$ of a suitable linear map from $\mathbb{R}^n\times\mathbb{R}^n$ into $\mathbb{R}^n$. So it is reasonable to ask whether the linear map that produces $\pi$ is uniquely determined by $\pi$. Otherwise stated, we want to know if the pair of $n\times n$

matrices $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)})$ that realizes $\pi$ in (2.1) is unique. This is answered in the following lemma.

LEMMA 2.2. *Assume* $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)})$ *is a pair of* $n \times n$ *matrices of a 2D Markov chain* $\mathscr{M}$ *with n states that realizes the transition map* (2.4). *Then, for any matrix* $\mathbf{M}$ *with all rows the same vector, the pair* $(\mathbf{A}^{(1)}+\mathbf{M}, \mathbf{A}^{(2)}-\mathbf{M})$ *realizes the same transition map. Vice versa, if* $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)})$ *and* $(\overline{\mathbf{A}}^{(1)}, \overline{\mathbf{A}}^{(2)})$ *realize* (2.4), *then there exists a matrix* $\mathbf{M}$ *with all rows the same vector such that*

$$\overline{\mathbf{A}}^{(1)} = \mathbf{A}^{(1)} + \mathbf{M}, \qquad \overline{\mathbf{A}}^{(2)} = \mathbf{A}^{(2)} - \mathbf{M}. \tag{2.5}$$

*Proof*. Let $\mathbf{v}$ be any probability vector and $\mathbf{M} \in \mathbb{R}^{n \times n}$ any matrix of the following form:

$$\mathbf{M} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n] = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \dots & \dots & \dots & \dots \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \end{bmatrix}. \tag{2.6}$$

It is easy to check that

$$\mathbf{vM} = [\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_n]$$

is independent of $\mathbf{v} \in X$. As a consequence, given any pair $\mathbf{x}(h, k+1)$, $\mathbf{x}(h+1, k)$ of probability vectors, the updated vector satisfies

$$\begin{aligned} \mathbf{x}(h+1, k+1) &= \mathbf{x}(h, k+1)\mathbf{A}^{(1)} + \mathbf{x}(h+1, k)\mathbf{A}^{(2)} \\ &= \mathbf{x}(h, k+1)[\mathbf{A}^{(1)} + \mathbf{M}] + \mathbf{x}(h+1, k)[\mathbf{A}^{(2)} - \mathbf{M}]. \end{aligned}$$

Therefore $(\mathbf{A}^{(1)}+\mathbf{M}, \mathbf{A}^{(2)}-\mathbf{M})$ and $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)})$ give rise to equivalent 2D Markov chains.

Vice versa, suppose that

$$\mathbf{x}(h, k+1)\mathbf{A}^{(1)} + \mathbf{x}(h+1, k)\mathbf{A}^{(2)} = \mathbf{x}(h, k+1)\overline{\mathbf{A}}^{(1)} + \mathbf{x}(h+1, k)\overline{\mathbf{A}}^{(2)} \tag{2.7}$$

holds for any pair of probability vectors $\mathbf{x}(h, k+1), \mathbf{x}(h+1, k)$. Letting

$$\begin{aligned} \mathbf{M} &= \mathbf{A}^{(1)} - \overline{\mathbf{A}}^{(1)}, \\ \mathbf{N} &= \mathbf{A}^{(2)} - \overline{\mathbf{A}}^{(2)}, \end{aligned} \tag{2.8}$$

and

$$\mathbf{x}(h,k+1)=\mathbf{e}_i=\begin{bmatrix}0 & 0 & \cdots & \underbrace{1}_{i\text{th place}} & \cdots & 0\end{bmatrix}, \qquad i=1,2,\ldots,n,$$

$$\mathbf{x}(h+1,k)=\mathbf{e}_j=\begin{bmatrix}0 & 0 & \cdots & \underbrace{1}_{j\text{th place}} & \cdots & 0\end{bmatrix}, \qquad j=1,2,\ldots,n,$$

from (2.7) we have

$$\mathbf{e}_i\mathbf{M}+\mathbf{e}_j\mathbf{N}=\mathbf{0}, \qquad i,j=1,2,\ldots,n. \tag{2.9}$$

This shows that $\mathbf{M}$ has the structure (2.6) and $\mathbf{M}=-\mathbf{N}$. ■

As a consequence of the above lemma, given a 2D Markov chain with $n$ states, there are infinitely many chains equivalent to it (i.e. that realize the same probability transition map). The equivalence is expressed by Equation (2.5), where $\mathbf{M}$ belongs to the space of matrices $n\times n$ with all rows the same vector. We are now in a position to prove that each equivalence class includes at least one 2D Markov chain represented by a convex combination of two stochastic matrices.

THEOREM 2.1. *A 2D Markov chain with n states can be represented as*

$$\mathbf{x}(h+1,k+1)=\mathbf{x}(h,k+1)a\mathbf{P}+\mathbf{x}(h+1,k)(1-a)\mathbf{Q} \tag{2.10}$$

*where* $\mathbf{P}$ *and* $\mathbf{Q}$ *are* $n\times n$ *stochastic matrices and* $0\leqslant a\leqslant 1$.

*Proof.* Suppose

$$\mathbf{x}(h+1,k+1)=\mathbf{x}(h,k+1)\mathbf{A}^{(1)}+\mathbf{x}(h+1,k)\mathbf{A}^{(2)} \tag{2.11}$$

is a 2D Markov chain. Let $\mathbf{M}$ denote a matrix with structure (2.6) and

$$\alpha_j:=-\min_{1\leqslant i\leqslant n} A_{ij}^{(1)}, \qquad j=1,2,\ldots,n.$$

By Lemma 2.2, the 2D Markov chains (2.11) and

$$\mathbf{x}(h+1,k+1)=\mathbf{x}(h,k+1)\overline{\mathbf{A}}^{(1)}+\mathbf{x}(h+1,k)\overline{\mathbf{A}}^{(2)} \tag{2.12}$$

with $\overline{\mathbf{A}}^{(1)} = \mathbf{A}^{(1)} + \mathbf{M}$, $\overline{\mathbf{A}}^{(2)} = \mathbf{A}^{(2)} - \mathbf{M}$ realize the same transition-probability map. By construction, $\overline{\mathbf{A}}^{(1)}$ is nonnegative and every column of it includes at least one zero element. Hence $\overline{\mathbf{A}}^{(2)}$ is nonnegative too. In fact, assume by contradiction that some element $\overline{A}^{(2)}_{ij}$ is negative, and consider any zero element in the $j$th column of $\overline{\mathbf{A}}^{(1)}$, say $\overline{A}^{(1)}_{hj} = 0$. Thus the $i$th entry of

$$\mathbf{e}_h \overline{\mathbf{A}}^{(1)} + \mathbf{e}_j \overline{\mathbf{A}}^{(2)} \tag{2.13}$$

is negative, and (2.13) cannot be a stochastic vector, because its $i$th entry is negative. Consequently $\overline{A}^{(2)}_{ij} \geqslant 0$, $i, j = 1, 2, \dots, n$.

Since for all pairs of probability vectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}$ the sum of the entries of $\mathbf{v}^{(1)}\overline{\mathbf{A}}^{(1)} + \mathbf{v}^{(2)}\overline{\mathbf{A}}^{(2)}$ is one, i.e.

$$\sum_{i=1}^{n} (\mathbf{v}^{(1)}\overline{\mathbf{A}}^{(1)})_i + \sum_{i=1}^{n} (\mathbf{v}^{(2)}\overline{\mathbf{A}}^{(2)})_i = 1, \tag{2.14}$$

we have

$$\sum_{i=1}^{n} (\mathbf{v}^{(1)}\overline{\mathbf{A}}^{(1)})_i = a \qquad \forall \mathbf{v}^{(1)} \in X,$$

$$\sum_{i=1}^{n} (\mathbf{v}^{(2)}\overline{\mathbf{A}}^{(2)})_i = 1 - a \qquad \forall \mathbf{v}^{(2)} \in X \tag{2.15}$$

for some $a \in \mathbb{R}$. Moreover, $a$ and $1-a$ are nonnegative because of the nonnegativity of $\overline{\mathbf{A}}^{(1)}$ and $\overline{\mathbf{A}}^{(2)}$, which amounts to saying that $a$ belongs to the interval $[0, 1]$.

If we define

$$\mathbf{P} = \begin{cases} \overline{\mathbf{A}}^{(1)}/a & \text{in case} \quad a \neq 0, \\ \mathbf{I}_n & \text{in case} \quad a = 0 \end{cases}$$

and

$$\mathbf{Q} = \begin{cases} \dfrac{\overline{\mathbf{A}}^{(2)}}{1-a} & \text{in case} \quad a \neq 0, \\ \mathbf{I}_n & \text{in case} \quad a = 0, \end{cases}$$

we easily see that both $\mathbf{P}$ and $\mathbf{Q}$ are stochastic matrices. Thus (2.10) is proved. ■

REMARK. The above theorem completely clarifies the class of dynamical models described by Equation (2.1). Actually we may visualize the process which moves from states $S_f$ at $(h,k+1)$ and $S_g$ at $(h+1,k)$ to some state at $(h+1,k+1)$ according to the following rules:

(1) The probability vectors $\mathbf{x}(h,k+1)$ and $\mathbf{x}(h+1,k)$ are thought of as giving the probabilities for the various possible starting states. Then an experiment in two stages takes place at $(h+1,k+1)$:

(2) The first stage of the experiment exhibits two possible outcomes, e.g. $\theta(h+1,k+1)=0$ and $\theta(h+1,k+1)=1$, with probabilities $a$ and $1-a$ respectively. The random variable $\theta(h+1,k+1)$ is independent of $\theta(l,m)$ for all $(l,m)\neq(h+1,k+1)$

(3) At the second stage a state transition occurs that uniquely depends on the state at $(h,k+1)$ if $\theta(h+1,k+1)=0$, and on the state at $(h+1,k)$ if $\theta(h+1,k+1)=1$. The process moves from $S_f$ at $(h,k+1)$ into $S_m$ with probability $P_{fm}$, and from $S_g$ at $(h+1,k)$ into $S_m$ with probability $Q_{gm}$.

In the sequel a chain in the form (2.10) will be called a *canonical 2D Markov chain* and will be denoted as $\mathscr{M}=(a,\mathbf{P},\mathbf{Q})$. This implies a slight abuse of language, since the equivalence classes need not include a single canonical chain, as shown by the following example.

EXAMPLE 2.1. The 2D system

$$\mathbf{x}(h+1,k+1)=\mathbf{x}(h,k+1)\begin{bmatrix}1/4 & 1/4\\ 1/6 & 2/6\end{bmatrix}+\mathbf{x}(h+1,k)\begin{bmatrix}1/2 & 0\\ 2/6 & 1/6\end{bmatrix}\tag{2.16}$$

is a canonical 2D Markov chain with two states. Indeed, its matrices can be rewritten as

$$\frac{1}{2}\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{3} & \frac{2}{3}\end{bmatrix}=a\mathbf{P},\qquad \frac{1}{2}\begin{bmatrix}1 & 0\\ \frac{2}{3} & \frac{1}{3}\end{bmatrix}=(1-a)\mathbf{Q}.$$

Computing the matrix $\mathbf{M}$ as in the proof of Theorem 2.1, we find

$$\mathbf{M}=\begin{bmatrix}-\frac{1}{6} & -\frac{1}{4}\\ -\frac{1}{6} & -\frac{1}{4}\end{bmatrix},$$

and the pair

$$a\mathbf{P} = \mathbf{M} = \frac{1}{12}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad (1-a)\mathbf{Q} - \mathbf{M} = \frac{11}{12}\begin{bmatrix} \frac{8}{11} & \frac{3}{11} \\ \frac{6}{11} & \frac{5}{11} \end{bmatrix} \tag{2.17}$$

gives a canonical 2D Markov chain equivalent to (2.16). Note that, assuming

$$\mathbf{M}' = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix},$$

we still obtain a 2D Markov chain equivalent to (2.16):

$$a\mathbf{P} = \mathbf{M}' = \begin{bmatrix} 10/12 & -9/12 \\ 14/12 & -8/12 \end{bmatrix}, \qquad (1-a)\mathbf{Q} - \mathbf{M}' = \begin{bmatrix} -1/2 & 1 \\ -4/6 & 7/6 \end{bmatrix}. \tag{2.18}$$

Clearly (2.18) is no longer canonical.

To conclude this section, we wish to investigate what matrix structures are allowed for 2D Markov chains when the dynamics of the probability vectors is one-dimensional. That is, we want to characterize the pairs $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)})$ that provide (canonical and noncanonical) 2D Markov chains equivalent to the following ones:

$$\mathbf{x}(h+1, k+1) = \mathbf{x}(h+1, k)\overline{\mathbf{A}}^{(2)} \tag{2.19}$$

or

$$\mathbf{x}(h+1, k=1) = \mathbf{x}(h, k+1)\overline{\mathbf{A}}^{(1)}. \tag{2.20}$$

Clearly, a 2D Markov chain equivalent to (2.19) or to (2.20) has matrices

$$\mathbf{A}^{(1)} = \mathbf{M}, \qquad \mathbf{A}^{(2)} = \overline{\mathbf{A}}^{(2)} - \mathbf{M} \tag{2.21}$$

or

$$\mathbf{A}^{(1)} = \overline{\mathbf{A}}^{(1)} - \mathbf{M}, \qquad \mathbf{A}^{(2)} = \mathbf{M}, \tag{2.22}$$

where $\mathbf{M}$ is an arbitrary $n \times n$ matrix with all rows the same vector.

The converse is also true. Indeed, a 2D Markov chain where $\mathbf{A}^{(1)}$ or $\mathbf{A}^{(2)}$ is a matrix with all rows the same vector is equivalent to (2.19) or to (2.20). Therefore (2.21) and (2.22) provide the most general structure of 2D Markov chains with one-dimensional dynamics.

If we concentrate our attention on canonical 2D Markov chains only, some restrictions on the structure of $\mathbf{M}$ are needed in Equation (2.21), to guarantee that $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ constitute a convex combination of stochastic matrices. First, requiring $\mathbf{A}^{(1)} = \mathbf{M}$ implies that in

$$\mathbf{M} = a \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} [\, p_1 \quad p_2 \quad \cdots \quad p_n \,], \tag{2.23}$$

$a$ must belong to the interval $[0,1]$ and $[p_1 \ p_2 \ \cdots \ p_n]$ must be a probability vector. Further restrictions on $\mathbf{M}$ depend on the requirement that $\mathbf{A}^{(2)}$ be a nonnegative matrix. For, if $\delta_j := \min_i \overline{A}^{(2)}_{ij}$, $j = 1, 2, \ldots, n$, denotes the minimum entry of each column of $\overline{\mathbf{A}}^{(2)}$, we must have

$$ap_j \leqslant \delta_j \tag{2.24}$$

and consequently

$$a \leqslant \sum_{j=1}^{n} \delta_j. \tag{2.25}$$

On the other hand, if $0 \leqslant a \leqslant \Sigma_j \delta_j$, there exists a probability vector that satisfies (2.24), and the corresponding matrix $\mathbf{M}$ provides, via (2.21), a canonical 2D Markov chain. Thus (2.19) does admit many equivalent canonical 2D Markov chains if and only if all entries of some column of $\overline{\mathbf{A}}^{(2)}$ are strictly positive. Obviously, the same result holds for $\mathbf{A}^{(1)}$ in Equation (2.20).

Remark. A stochastic $n \times n$ matrix $\mathbf{A}^{(2)}$ with a strictly positive column exhibits strong spectral properties. For, suppose $A^{(2)}_{i1} > 0$, $i = 1, 2, \ldots, n$. Since the corresponding 1D Markov chain with $n$ states is allowed to jump from every state to the state $S_1$ in one step, $S_1$ belongs to the unique ergodic class [5] of the chain. We order the states with $S_1$ first, followed by all those associated with the ergodic class and finally by those associated with transient classes. If the states are ordered this way, the transition matrix can be written in partitioned form

$$\mathbf{A}^{(2)} = \begin{bmatrix} \mathbf{E} & \mathbf{0} \\ \mathbb{R} & \mathbf{T} \end{bmatrix}. \tag{2.26}$$

The transition probabilities within the ergodic class are represented by the submatrix $\mathbf{E}$, whose first column is strictly positive. This inhibits the ergodic class from being periodic. Therefore, if (2.19) is equivalent to a canonical 2D Markov chain $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ with $0 < a < 1$, then $\overline{\mathbf{A}}^{(2)}$ has the eigenvalue 1, which is a simple root of the characteristic equation [6]. The magnitudes of the other eigenvalues are less than 1.

## 3. 2D CHARACTERISTIC POLYNOMIAL

When viewed in terms of its probability vectors evolving in $\mathbb{Z} \times \mathbb{Z}$, a 2D Markov chain is a 2D system whose system matrices are given by a convex combination of a pair of stochastic matrices. Thus it is expected that the strong spectral properties of stochastic matrices will play a central role in the theory of 2D Markov chains. Indeed this is true, as we shall see when considering the long-term distribution of states and the existence of stable probability configurations.

To analyze these asymptotic phenomena, it is convenient to introduce the so-called 2D characteristic polynomial and to study in some detail the algebraic variety of its zero set. It is well known that this topic forms the framework for much of the internal stability analysis of general 2D systems [2, 7]. Here, however, the peculiar structure of $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ induces some *a priori* constraints on the polynomial variety, which will be of use in the next section.

Consider a canonical 2D Markov chain with $n$ states $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$, given by Equation (2.10). The polynomial in two indeterminates

$$\Delta(z_1, z_2) = \det[\mathbf{I} - a z_1 \mathbf{P} - (1-a) z_2 \mathbf{Q}] \tag{3.1}$$

is called the *characteristic polynomial* of $\mathscr{M}$, and the solutions of the corresponding equation

$$\Delta(z_1, z_2) = 0 \tag{3.2}$$

constitute the variety $\mathscr{V}(\Delta)$ of the chain.

It is possible to derive a simple set of conditions on the structure of $\Delta(z_1, z_2)$, that depend on the stochastic nature of $\mathbf{P}$ and $\mathbf{Q}$. Let us consider the subspace of $\mathbb{R}^n$ consisting of all row vectors whose entries sum to zero:

$$N := \left\{ \mathbf{v} = [\, v_1 \quad v_2 \quad \cdots \quad v_n \,] \in \mathbb{R}^n \,\middle|\, \sum_{i=1}^{n} v_i = 0 \right\}. \tag{3.3}$$

$N$ is an invariant subspace [8] relative to the matrices $\mathbf{P}$ and $\mathbf{Q}$. For, given any $\mathbf{v} \in N$, we have

$$\sum_{j=1}^{n} (\mathbf{vP})_j = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} v_i P_{ij} \right) = \sum_{i=1}^{n} v_i \sum_{j=1}^{n} P_{ij} = \sum_{i=1}^{n} v_i = 0$$

and analogously

$$\sum_{j=1}^{n} (\mathbf{vQ})_j = 0.$$

Let $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n)$ be a basis for $\mathbb{R}^n$ such that $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{n-1})$ is a basis for $N$ and $\mathbf{r}_n$ a probability vector. After introducing the nonsingular matrix

$$\mathbf{T} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_n \end{bmatrix}, \tag{3.4}$$

any vector $\mathbf{v} \in \mathbb{R}^n$ will be represented by the $n$-tuple $\hat{\mathbf{v}} = [\hat{v}_1 \ \hat{v}_2 \ \cdots \ \hat{v}_n] = \mathbf{vT}^{-1}$ in the new basis. Moreover, the linear transformations represented by $\mathbf{P}$ and $\mathbf{Q}$ in the standard basis will be represented by

$$\hat{\mathbf{P}} = \mathbf{TPT}^{-1} = \begin{bmatrix} \hat{\mathbf{P}}_{11} & \mathbf{0} \\ \hat{\mathbf{P}}_{21} & \hat{P}_{22} \end{bmatrix} \tag{3.5}$$

and

$$\hat{\mathbf{Q}} = \mathbf{TQT}^{-1} = \begin{bmatrix} \hat{\mathbf{Q}}_{11} & \mathbf{0} \\ \hat{\mathbf{Q}}_{21} & \hat{Q}_{22} \end{bmatrix} \tag{3.6}$$

in the new basis.

The components $\hat{w}_1, \hat{w}_2, \dots, \hat{w}_n$ of the probability vector $\mathbf{r}_n\mathbf{P}$ with respect to the new basis

$$\mathbf{r}_n\mathbf{P} = \hat{w}_1\mathbf{r}_1 + \hat{w}_2\mathbf{r}_2 + \dots + \hat{w}_n\mathbf{r}_n$$

are the entries of $[\hat{\mathbf{P}}_{21} \quad \hat{P}_{22}]$. Since we have

$$\begin{aligned} 1 = \sum_{i=1}^{n} (\mathbf{r}_n\mathbf{P})_i &= \sum_{i=1}^{n}\sum_{j=1}^{n-1} (\hat{w}_j\mathbf{r}_j)_i + \sum_{i=1}^{n} (\hat{w}_n\mathbf{r}_n) \\ &= \sum_{j=1}^{n-1} \hat{w}_j \sum_{i=1}^{n} r_{ji} + \hat{w}_n \\ &= \hat{w}_n, \end{aligned}$$

we see that $\hat{P}_{22}$ and, by the same argument, $\hat{Q}_{22}$ are equal to 1. As a direct consequence of the block triangular structure of $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$, the characteristic polynomial of $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ factorizes as

$$\Delta(z_1, z_2) = [1 - az_1 - (1-a)z_2]\det[\mathbf{I}_{n-1} - az_1\hat{\mathbf{P}}_{11} - (1-a)z_2\hat{\mathbf{Q}}_{11}]. \tag{3.7}$$

It must be emphasized that the characteristic polynomial of a 2D Markov chain is not invariant under the equivalence (2.5) induced by matrices (2.6). Actually, any matrix

$$\mathbf{M} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_n]$$

reduces by similarity to

$$\hat{\mathbf{M}} = \mathbf{TMT}^{-1} = \begin{bmatrix} 0 & 0 \\ * & \mu \end{bmatrix}$$

with $\mu = \sum_{i=1}^{n} \mu_i$. Therefore the matrices $a\mathbf{P} + \mathbf{M}$, $(1-a)\mathbf{P} - \mathbf{M}$ of any (not necessarily canonical) chain equivalent to $\mathscr{M}$ are similar to

$$a\hat{\mathbf{P}}+\hat{\mathbf{M}}=\begin{bmatrix} a\hat{\mathbf{P}}_{11} & 0 \\ * & a+\mu \end{bmatrix},$$

$$(1-a)\hat{\mathbf{Q}}-\hat{\mathbf{M}}=\begin{bmatrix} (1-a)\hat{\mathbf{Q}}_{11} & 0 \\ * & 1-(a+\mu) \end{bmatrix},$$

and the corresponding characteristic polynomial is

$$\Delta_{\mathbf{M}}(z_1,z_2)=[1-(a+\mu)z_1-(1-a-\mu)z_2]$$
$$\times\det\left[\mathbf{I}_{n-1}-az_1\hat{\mathbf{P}}_{11}-(1-a)z_2\hat{\mathbf{Q}}_{11}\right].$$

The results obtained so far are summarized in the following

THEOREM 3.1. *The characteristic polynomial of a 2D Markov chain with n states factorizes into the product of a first-order polynomial*

$$h_1(z_1,z_2)=1-az_1-(1-a)z_2 \tag{3.8}$$

*and a polynomial $h_2(z_1,z_2)$ of degree not greater than $n-1$. While $h_2$ is invariant under the 2D chain equivalence (2.5), $h_1$ is not, and its orbit is obtained by varying the parameter a arbitrarily over the real numbers. In canonical 2D Markov chains, $0\leqslant a\leqslant 1$.*

EXAMPLE 3.1. Consider once again the canonical 2D Markov chain (2.16). Its characteristic polynomial is

$$\Delta(z_1,z_2)=\left(1-\tfrac{1}{2}z_1-\tfrac{1}{2}z_2\right)\left(1-\tfrac{1}{12}z_1-\tfrac{1}{6}z_2\right).$$

The equivalent noncanonical Markov chain (2.18) has exactly the same characteristic polynomial. This shows that the condition $0\leqslant a\leqslant 1$ in $h_1(z_1,z_2)$ is necessary but not sufficient to guarantee a 2D Markov chain to be canonical.

Finally, the characteristic polynomial of (2.17) is

$$\left(1-\tfrac{1}{12}z_1-\tfrac{11}{12}z_2\right)\left(1-\tfrac{1}{12}z_1-\tfrac{1}{6}z_2\right).$$

Theorem 3.1 provides a first insight into the structure of the variety of a 2D Markov chain. Actually, if we consider the variety of the first-order polynomial $h_1(z_1, z_2)$ in (3.8), we see immediately that

(1) $(1,1) \in \mathscr{V}(h_1) \subseteq \mathscr{V}(\Delta)$;

(2) if $0 \leqslant a \leqslant 1$, then $(1,1)$ is the only point where $\mathscr{V}(h_1)$ intersects the unit closed polydisk

$$\mathscr{P}_1 = \{(z_1, z_2) \,|\, |z_1| \leqslant 1,\ |z_2| \leqslant 1\}. \tag{3.9}$$

The next theorem below shows that, in the case where a canonical 2D Markov chain is nontrivial (i.e. $0 < a < 1$), the intersections between the complete variety of the chain $\mathscr{V}(\Delta)$ and the unit polydisk are restricted to the distinguished boundary

$$\mathscr{T}_1 = \{(z_1, z_2) \,|\, |z_1| = |z_2| = 1\} \tag{3.10}$$

of the unit polydisk.

THEOREM 3.2. *Assume that in a canonical 2D Markov chain* $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ *both* $a$ *and* $1-a$ *are different from zero. Then the variety* $\mathscr{V}(\Delta)$ *does not intersect the unit closed polydisk* $\mathscr{P}_1$ *except at* $(1,1)$ *and, possibly, at some other points of its distinguished boundary* $\mathscr{T}_1$.

*Proof.* Assume, by contradiction, that $\mathscr{V}(\Delta)$ intersects $\mathscr{P}_1 \setminus \mathscr{T}_1$ at $(\rho_1 e^{i\omega_1}, \rho_2 e^{i\omega_2})$. Thus there exists a nonzero $\mathbf{v} \in \mathbf{C}^n$ such that

$$\mathbf{v} = \mathbf{v} a \rho_1 e^{i\omega_1} \mathbf{P} + \mathbf{v}(1-a)\rho_2 e^{i\omega_2} \mathbf{Q},$$

and consequently

$$e^{-i\omega_1}\mathbf{v} = \mathbf{v}\left[a\rho_1 \mathbf{P} + (1-a)\rho_2 e^{i\omega}\mathbf{Q}\right], \tag{3.11}$$

where

$$\omega = \omega_2 - \omega_1.$$

It is convenient to use the polar representation for the entries of $\mathbf{v}$:

$$\mathbf{v} = \begin{bmatrix} p_1 e^{i\beta_1} & p_2 e^{i\beta_2} & \cdots & p_n e^{i\beta_n} \end{bmatrix} \tag{3.12}$$

and normalize $\mathbf{v}$ so as to have

$$\sum_{h=1}^{n} p_h = \|\mathbf{v}\|_1 = 1. \tag{3.13}$$

Let $\mathbf{r}_h$ denote the $h$th row of $a\rho_1\mathbf{P}+(1-a)\rho_2 e^{i\omega}\mathbf{Q}$, $h=1,2,\ldots,n$, and rewrite (3.11) as

$$e^{-i\omega_1}\mathbf{v} = p_1 e^{i\beta_1}\mathbf{r}_1 + p_2 e^{i\beta_2}\mathbf{r}_2 + \cdots + p_n e^{i\beta_n}\mathbf{r}_n. \tag{3.14}$$

Computing the $l_1$ norm of $\mathbf{r}_h$ gives

$$\begin{aligned}\|\mathbf{r}_h\|_1 &= \sum_{k=1}^{n}\left| a\rho_1 \sum_{k=1}^{n} P_{hk} + (1-a)\rho_2 e^{i\omega} \sum_{k=1}^{n} Q_{hk}\right| \\ &\leqslant a\rho_1 \sum_{k=1}^{n} P_{hk} + (1-a)\rho_2 \sum_{k=1}^{n} Q_{hk} \qquad (h=1,2,\ldots,n) \\ &= a\rho_1 + (1-a)\rho_2.\end{aligned}$$

Since $0<a<1$, it is clear that all vectors $\mathbf{r}_h$ have an $l_1$ norm less than 1 and therefore, in view of (3.14),

$$\|\mathbf{v}\|_1 = \|e^{-i\omega_1}\mathbf{v}\|_1 \leqslant \sum_{h=1}^{n} p_h \|\mathbf{r}_h\|_1 < \sum_{h=1}^{n} p_h = 1,$$

which contradicts (3.13). ∎

As an immediate consequence of Theorems 3.1 and 3.2, we have the following

Corollary 3.1. *The variety $\mathscr{V}(h_1)$ of a canonical 2D Markov chain with $0<a<1$ does not intersect $\mathscr{P}_1 \setminus \mathscr{T}_1$.*

## 4. ASYMPTOTIC BEHAVIOR

As pointed our earlier, for certain types of 2D Markov chains there exists a unique limiting probability vector, independent of the distribution of the probability vectors $\mathbf{x}(h, -h)$ on the separation set $\mathscr{C}_0$. This class of chains, which can be regarded as the 2D analogue of 1D Markov chains with a single aperiodic class, has a deep but intuitive body of theory. The purpose of this section is to present a fairly simple criterion for identifying these chains, based on the structure of their characteristic polynomial.

Obviously, the case where 2D chains exhibit a one-dimensional dynamics is already solved using the standard 1D theory. So, in the following developments we shall consider only canonical 2D Markov chains $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ with $0 < a < 1$. Without loss of generality, we may also assume that the states of the 1D chain associated with the stochastic matrix

$$\mathbf{A} = a\mathbf{P} + (1-a)\mathbf{Q}$$

have been permuted so that all the ergodic states are listed before the transient states. In other words, without loss of generality, we will assume from now on that the above matrix is block triangular:

$$\mathbf{A} = \begin{bmatrix} \mathbf{E} & \mathbf{0} \\ \mathbb{R} & \mathbf{T} \end{bmatrix}, \tag{4.1}$$

where $\mathbf{E}$ and $\mathbf{T}$ are a stochastic and a substochastic matrix respectively, representing the transition probabilities within the ergodic classes and the transition probabilities among the transient states of a 1D Markov chain. Clearly, the partition (4.1) carries over to $\mathbf{P}$ and $\mathbf{Q}$, which will be written as follows:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{0} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{0} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix}. \tag{4.2}$$

Definition 4.1. Let $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ be a 2D Markov chain, and $\mathscr{X}_0$ a sequence of initial probability vectors. A probability vector $\mathbf{w} \in X$ is a *limiting probability vector* (LPV) of $\mathscr{X}_0$ if

$$\lim_{h+k \to +\infty} \mathbf{x}(h, k) = \mathbf{w}. \tag{4.3}$$

If (4.3) holds for all sequences $\mathscr{X}_0$ of initial probability vectors, $\mathbf{w}$ is termed a *global limiting probability vector* (GLPV).

As a direct consequence of the above definition, if $\mathbf{w}$ is a LPV of some sequence $\mathscr{X}_0$, it is also a LPV of the sequence

$$\hat{\mathscr{X}}_0 = \{\mathbf{x}(h, -h) = \mathbf{w},\ h \in \mathbb{Z}\}$$

We therefore have the following equivalent

Definition 4.1′. Let $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ be a 2D Markov chain. A probability vector $\mathbf{w} \in X$ is a LPV if

$$\mathbf{w} = \mathbf{w}[a\mathbf{P} + (1-a)\mathbf{Q}]. \tag{4.4}$$

The general strategy in studying the existence of a GLPV is to first derive some constraints on the values of its entries. Then one shows that the variety $\mathscr{V}(\Delta)$ of the characteristic polynomial must be regular at $(1,1)$ and, by a perturbation argument, cannot intersect the distinguished boundary $\mathscr{T}_1$ except at $(1,1)$. The above constraints on $\mathscr{V}(\Delta)$ are finally converted into sufficient conditions for $\mathscr{M}$ having a GLPV.

Lemma 4.1. *Let* $\mathbf{w}$ *be a LPV of* $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$, *and assume that in* (4.1) *the matrix* $\mathbf{T}$ *has dimension* $r \times r$. *Then the last* $r$ *entries of* $\mathbf{w}$ *are zero.*

*Proof.* Partition $\mathbf{w}$ conformably with the block triangular structure of (4.1):

$$\mathbf{w} = [\mathbf{w}_1 \quad \mathbf{w}_2] \tag{4.5}$$

Then, by Definition 1′, the last $r$ entries of $\mathbf{w}$ satisfy the following equation:

$$\mathbf{w}_2 = \mathbf{w}_2\mathbf{T}. \tag{4.6}$$

Since the spectral radius of $\mathbf{T}$ is less than 1, $\mathbf{w}_2 = 0$ is the unique solution of Equation (4.6). ∎

We recall that a stochastic matrix $\mathbf{C}$ and the corresponding 1D Markov chain are fully regular if $\mathbf{C}$ has no characteristic values of modulus 1 other than 1 itself and 1 is a simple root of the characteristic equation of $\mathbf{C}$ [6]. In this case the Markov chain consists of a single ergodic aperiodic class.

LEMMA 4.2. *Suppose that* $\mathscr{M}=(a,\mathbf{P},\mathbf{Q})$ *admits a GLPV* $\mathbf{w}$. *Then in* (4.1) *the matrix* $\mathbf{E}$ *is fully regular.*

*Proof.* Assume by contradiction that the ergodic states of (4.1) are partitioned in at least two communication classes. Then, possibly after a permutation of the ergodic states, $\mathbf{E}$ reduces to the following form:

$$\mathbf{E}=\operatorname{diag}(\mathbf{E}_1,\mathbf{E}_2,\dots,\mathbf{E}_t),\qquad t\geqslant 2,$$

where $\mathbf{E}_h$ are irreducible stochastic matrices of dimension $\nu_h\times\nu_h$. Let $\mathbf{p}_h\in\mathbb{R}^{\nu_h}$ denote the unique probability vector such that $\mathbf{p}_h\mathbf{E}_h=\mathbf{p}_h$, $h=1,2$. Thus both

$$\Big[\underbrace{\mathbf{p}_1}_{\nu_1}\Big|\underbrace{0\ \cdots\ 0}_{\nu_2}\Big|\underbrace{0\ \cdots\ 0}_{n-\nu_1-\nu_2}\Big]$$

and

$$\Big[\underbrace{0\ \cdots\ 0}_{\nu_1}\Big|\underbrace{\mathbf{p}_2}_{\nu_2}\Big|\underbrace{0\ \cdots\ 0}_{n-\nu_1-\nu_2}\Big]$$

are left eigenvectors of $\mathbf{A}$ corresponding to the eigenvalue $\lambda_0=1$, and, by Definition 1′, $\mathscr{M}$ would have two distinct LPVs, which contradicts the GLPV assumption. Thus the ergodic states of (4.1) constitute a single ergodic class.

Suppose now that the ergodic class is periodic, with period $\mu>1$. Then there exists a probability vector $\mathbf{p}\in\mathbb{R}^{n-r}$ such that

$$\mathbf{p},\mathbf{pE},\dots,\mathbf{pE}^{\mu-1}$$

are different each other. It is quite easy to check that, if the sequence of initial probability vectors is

$$\mathscr{X}_0=\{\mathbf{x}(h,-h)=[\mathbf{p}\,|\,\mathbf{0}],\ h\in\mathbb{Z}\},$$

then

$$\mathscr{X}_k=\mathscr{X}_h\quad\text{iff}\quad k\equiv h\ (\operatorname{mod}\ \mu).$$

This amounts to saying that $\mathscr{M}$ undergoes a periodic evolution and rules out once more the possibility of having a GLPV. ■

The following theorem is a direct consequence of the above lemmas.

THEOREM 4.1. *Let* $\mathbf{w}$ *be a GLPV of* $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$, *and assume that in* (4.1) $\mathbf{T}$ *has dimension* $r \times r$. *Then the entries of* $\mathbf{w}_1$ *in*

$$\mathbf{w} = \left[\mathbf{w}_1 \;\middle|\; \underbrace{0 \;\cdots\; 0}_{r}\right]$$

*are strictly positive.*

*Proof.* Consider a constant sequence of initial probability vectors of the following form:

$$\mathscr{X}_0 = \{\mathbf{x}(h, -h) = [\mathbf{v}_1 \,|\, \mathbf{0}],\, h \in \mathbb{Z}\}.$$

Then $\mathscr{X}_k = \{\mathbf{x}(h, -h+k) = [\mathbf{v}_1 \mathbf{E}^k \,|\, \mathbf{0}],\, h \in \mathbb{Z}\}$ is a constant sequence too. By the GLPV assumption, $[\mathbf{v}_1 \mathbf{E}^k \,|\, \mathbf{0}]$ must converge to $\mathbf{w}$ as $k$ increases, or equivalently,

$$\mathbf{w}_1 = \lim_{k \to +\infty} \mathbf{v}_1 \mathbf{E}^k.$$

Since $\mathbf{E}$ is fully regular, $\mathbf{E}^k$ has no zero entries for large values of $k$, and $\mathbf{w}_1$ is strictly positive. ■

If the characteristic polynomial of a chain $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ has repeated roots at $(1,1)$, it is impossible to find a GLPV. The same happens if some roots belong to $\mathscr{T}_1 \setminus \{(1,1)\}$. To discuss the first property, we shall need the following technical lemma:

LEMMA 4.3. *Consider the factorization of* $\Delta(z_1, z_2)$ *given in* (3.7). *The following are equivalent:*

(1) $\lambda_0 = 1$ *is a multiple eigenvalue of* $\mathbf{A} = a\mathbf{P} + (1-a)\mathbf{Q}$;
(2) *when evaluated at* $(1,1)$, $\partial\Delta / \partial z_1$ *is zero*;
(3) *when evaluated at* $(1,1)$, $\partial\Delta / \partial z_2$ *is zero.*

*Proof.* Evaluating at $(1,1)$ the partial derivatives of (3.7), one gets

$$\left.\frac{\partial \Delta}{\partial z_1}\right|_{(1,1)} = -a \det\left[\mathbf{I} - a\hat{\mathbf{P}}_{11} - (1-a)\hat{\mathbf{Q}}_{11}\right], \tag{4.7}$$

$$\left.\frac{\partial \Delta}{\partial z_2}\right|_{(1,1)} = -(1-a) \det\left[\mathbf{I} - a\hat{\mathbf{P}}_{11} - (1-a)\hat{\mathbf{Q}}_{11}\right]. \tag{4.8}$$

Since $0 < a < 1$, the condition $\partial \Delta / \partial z_1|_{(1,1)} = 0$ is equivalent to

$$\det\left[\mathbf{I} - a\hat{\mathbf{P}}_{11} - (1-a)\hat{\mathbf{Q}}_{11}\right] = 0,$$

which in turn is equivalent to assuming that $\lambda_0 = 0$ has multiplicity greater than 1. Thus (1) $\Leftrightarrow$ (2) and, by a similar argument, (1) $\Leftrightarrow$ (3). ∎

THEOREM 4.2. *Let $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ have a GLPV. Then the variety $\mathscr{V}(\Delta)$ of its characteristic polynomial is regular at $(1,1)$.*

*Proof.* Suppose $(1,1)$ be a singular point of $\mathscr{V}(\Delta)$. By Lemma 4.3, $\lambda_0 = 1$ is a multiple eigenvalue of $\mathbf{A}$, and consequently $\mathbf{E}$ cannot be fully regular. This would contradict the existence of a GLPV. ∎

We consider now the possibility that the variety $\mathscr{V}(\Delta)$ and the distinguished boundary $\mathscr{T}_1$ may have intersections other than $(1,1)$ or, equivalently, the matrix $\mathbf{I} - az_1\mathbf{P} - (1-a)\mathbf{Q}$ may not be full rank at $(e^{i\omega_1}, e^{i\omega_2}) \neq (1,1)$.

LEMMA 4.4. *Let $\mathbf{P}_{22}$ and $\mathbf{Q}_{22}$ be as in the partition (4.2) of $\mathbf{P}$ and $\mathbf{Q}$. Then the polynomial $\det[\mathbf{I} - az_1\mathbf{P}_{22} - (1-a)z_2\mathbf{Q}_{22}]$ is devoid of zeros in the unit closed polydisk.*

*Proof.* Given a complex-valued matrix $\mathbf{C}$, we denote by $\operatorname{mod}\mathbf{C}$ the matrix which arises from $\mathbf{C}$ when all the elements are replaced by their moduli. It is easy to see that, for $(z_1, z_2) \in \mathscr{P}_1$,

$$\operatorname{mod}\left[az_1\mathbf{P}_{22} + (1-a)z_2\mathbf{Q}_{22}\right] \leqslant a\mathbf{P}_{22} + (1-a)\mathbf{Q}_{22}. \tag{4.9}$$

This implies that, for any eigenvalue $\gamma$ of $az_1\mathbf{P}_{22} + (1-a)z_2\mathbf{Q}_{22}$,

$$|\gamma| \leqslant R < 1, \tag{4.10}$$

where $R$ is the spectral radius of $a\mathbf{P}_{22} + (1-a)\mathbf{Q}_{22}$. Actually, if $a\mathbf{P}_{22} + (1-a)\mathbf{Q}_{22}$ is irreducible, (4.10) is a classical result needed in the proof of the Frobenius theorem [6]. The generalization to arbitrary nonnegative matrices is obtained by a limiting process, since $a\mathbf{P}_{22} + (1-a)\mathbf{Q}_{22}$ can be represented as the limit of a decreasing sequence of positive (and thus irreducible) matrices.

We therefore have that $\gamma\mathbf{I} - az_1\mathbf{P}_{22} - (1-a)z_2\mathbf{Q}_{22}$ cannot be singular if $(z_1, z_2)$ belongs to $\mathscr{P}_1$ and $|\gamma| \geqslant 1$. This proves the lemma. ■

LEMMA 4.5. *Suppose that* $\mathbf{I} - az_1\mathbf{P} - (1-a)z_2\mathbf{Q}$ *is not full rank at* $(e^{i\omega_1}, e^{i\omega_2}) \neq (1,1)$. *If* $\mathbf{v} = [v_1 \ v_2 \ \cdots \ v_n] \in \mathbb{C}^n$ *satisfies*

$$\mathbf{v}\left[\mathbf{I} - ae^{i\omega_1}\mathbf{P} - (1-a)e^{i\omega_2}\mathbf{Q}\right] = \mathbf{0}, \tag{4.11}$$

*then its entries sum to zero*:

$$\sum_{k=1}^{n} v_k = 0. \tag{4.12}$$

*Proof*. Letting $\omega := \omega_1 - \omega_2$ and denoting by $\mathbf{r}_h$, $h = 1, 2, \ldots, n$, the $h$th row of $a\mathbf{P} + (1-a)\mathbf{Q}e^{i\omega}$, we rewrite (4.11) as

$$e^{-i\omega_1}\mathbf{v} = \sum_{h=1}^{n} v_h \mathbf{r}_h. \tag{4.13}$$

Summing the entries of the row vectors on both sides of (4.3), one gets

$$\begin{aligned} e^{-i\omega_1}\sum_{k=1}^{n} v_h &= \sum_{h=1}^{n} v_h \sum_{k=1}^{n}\left[aP_{hk} + (1-a)e^{i\omega}Q_{hk}\right] \\ &= \sum_{h=1}^{n} v_h\left[a + (1-a)e^{i\omega}\right], \end{aligned}$$

which implies

$$\sum_{h=1}^{n} \{e^{-i\omega_1} - [a + (1-a)e^{i\omega}]\} = 0.$$

It follows that $\Sigma_{h=1}^{n} v_h = 0$, in view of the fact that, by assumption, either $\omega_1 \not\equiv 0$ or $\omega \not\equiv 0 \bmod 2\pi$. ■

In view of Lemmas 4.4 and 4.5, our original assumption on the existence of an intersection between $\mathscr{V}(\Delta)$ and $\mathscr{T}_1 \setminus \{(1,1)\}$ can be restated as follows: there exists a complex-valued nonzero vector

$$\mathbf{v} = [v_1 \quad v_2 \quad \cdots \quad v_{n-r}] \qquad \text{with} \quad \sum_{h=1}^{n-r} v_h = 0$$

that satisfies the following equation:

$$\mathbf{v}[\mathbf{I} - ae^{i\omega_1}\mathbf{P}_{11} - (1-a)e^{i\omega_2}\mathbf{Q}_{11}] = \mathbf{0}. \tag{4.14}$$

If we partition the probability vectors conformably with the block structure of (4.2),

$$\mathbf{x}(h,k) = [\mathbf{x}_1(h,k) \quad \mathbf{x}_2(h,k)],$$

and assume that the initial probability vectors satisfy $\mathbf{x}_2(h,-h) = \mathbf{0}$, $h \in \mathbb{Z}$, then the first $n-r$ entries of $\mathbf{x}(\cdot,\cdot)$ evolve according to the equation of a 2D Markov chain with $n-r$ states,

$$\mathbf{x}_1(h+1,k+1) = \mathbf{x}_1(h,k+1)a\mathbf{P}_{11} + \mathbf{x}(h+1,k)(1-a)\mathbf{Q}_{11}. \tag{4.15}$$

Suppose, for the moment, that in (4.15) all $\mathbf{x}_1$'s are allowed to be complex-valued vectors, and consider the following sequence:

$$\tilde{\mathscr{X}}_0 = \{\mathbf{x}_1(h,-h) = \mathbf{v}e^{i\omega h}, h \in \mathbb{Z}\} \tag{4.16}$$

with $\omega = \omega_2 - \omega_1$. It is clear that the updating equation (4.15) produces at $(h,k)$, with $h+k \geqslant 0$, a vector $\mathbf{x}_1(h,k) = \mathbf{v}e^{-i\omega_1 h - i\omega_2 k}$, and consequently the vector sequence $\tilde{\mathscr{X}}_m$ on the separation set $\mathscr{C}_m = \{(h,k) \mid h+k = m\}$ is given by

$$\tilde{\mathscr{X}}_m = \{\mathrm{x}_1(h, -h+m) = \mathrm{v}e^{i\omega h - i\omega_2 n},\ h \in \mathbb{Z}\} \tag{4.17}$$

Put another way, it is possible to recover $\tilde{\mathscr{X}}_m$ from $\tilde{\mathscr{X}}_0$ via multiplication by $e^{-i\omega_2 m}$.

When v is expressed in polar form

$$\mathrm{v} = \left[ p_1 e^{i\beta_1} \quad p_2 e^{i\beta_2} \quad \cdots \quad p_{n-r} e^{i\beta_{n-r}} \right],$$

the sequence (4.16) breaks apart into a real and an imaginary sequence:

$$\tilde{\mathscr{X}}_0 = \tilde{\mathscr{X}}_0^R + i\tilde{\mathscr{X}}_0^I,$$

with

$$\tilde{\mathscr{X}}_0^R = \{[p_1 \cos(\beta_1 + h\omega) \quad p_2 \cos(\beta_2 + h\omega) \quad \cdots \quad p_{n-r} \cos(\beta_{n-r} + h\omega)],\ h \in \mathbb{Z}\},$$

$$\tilde{\mathscr{X}}_0^I = \{[p_1 \sin(\beta_1 + h\omega) \quad p_2 \sin(\beta_2 + h\omega) \quad \cdots \quad p_{n-r} \sin(\beta_{n-r} + h\omega)],\ h \in \mathbb{Z}\}.$$

Since the transition matrices $a\mathbf{P}_{11}$ and $(1-a)\mathbf{Q}_{11}$ are real-valued, assuming $\tilde{\mathscr{X}}_0^R$ or $\tilde{\mathscr{X}}_0^I$ as initial conditions will produce separately.

$$\tilde{\mathscr{X}}_m^R = \{[p_1 \cos(\beta_1 + h\omega - m\omega_2) \quad p_2 \cos(\beta_2 + h\omega - m\omega_2) \quad \cdots \quad \ldots p_{n-r} \cos(\beta_{n-r} + h\omega - m\omega_2)],\quad h \in \mathbb{Z}\},$$

$$\tilde{\mathscr{X}}_0^I = \{[p_1 \sin(\beta_1 + h\omega - m\omega_2) \quad p_2 \sin(\beta_2 + h\omega - m\omega_2) \ldots \quad \ldots p_{n-r} \sin(\beta_{n-r} + h\omega - m\omega_2)],\quad h \in \mathbb{Z}\}.$$

Owing to the assumption $\mathrm{v} \neq \mathbf{0}$, $\tilde{\mathscr{X}}_0^R$ and $\tilde{\mathscr{X}}_0^I$ cannot be simultaneously zero. Furthermore, the property $\Sigma_h v_h = 0$ implies that the entries of every real vector of the sequences $\mathscr{X}_m^R$ and $\tilde{\mathscr{X}}_m^R$ sum to zero.

Suppose now we start the chain from $\tilde{\mathscr{X}}_0^R \neq 0$. Then the sequences $\tilde{\mathscr{X}}_m^R$ cannot converge to zero as $m$ goes to infinity. Actually, if $\omega_2/2\pi$ is rational, the sequences $\tilde{\mathscr{X}}_m^R$ vary periodically with $m$; if not, there are sequences $\tilde{\mathscr{X}}_m^R$ arbitrarily close to $\tilde{\mathscr{X}}_0^R$ for arbitrarily large values of $m$.

The above discussion is summarized in the following lemma

LEMMA 4.6. *Let* $\det[\mathbf{I} - az_1\mathbf{P} - (1-a)z_2\mathbf{Q}] = 0$ *at* $(z_1, z_2) = (e^{i\omega_1}, e^{i\omega_2}) \neq (1,1)$. *Then there exists a nonzero sequence of real vectors*

$$\mathscr{X}_0 = \left\{ \mathbf{x}(h, -h) = \Big[ \underbrace{\mathbf{x}_1}_{n-r} \Big| \mathbf{0} \Big], h \in \mathbb{Z} \right\} \tag{4.18}$$

*and two positive real numbers* $l \leqslant L$ *with the following properties: the vectors we obtain from* $\mathscr{X}_0$ *according to* (2.10) *satisfy*

(1) $x_j(h,k) = 0$, $n - r < j \leqslant n$,
(2) $\Sigma_{j=1}^n x_j(h,k) = 0$,
(3) $\|\mathscr{X}_m\| = \sup_{h \in \mathbb{Z}} \|x(h, -h+m)\|_\infty \in [l, L]$.

We are now in a position to prove the main result of this section.

THEOREM 4.3. *Let* $\mathscr{M} = (a, \mathbf{P}, \mathbf{Q})$ *be a 2D Markov chain. Then* $\mathscr{M}$ *admits a GLPV if and only if* $(1,1)$ *is a regular point of* $\mathscr{V}(\Delta)$ *and is the unique intersection of* $\mathscr{V}(\Delta)$ *with the distinguished boundary* $\mathscr{T}_1$.

*Proof*. To prove the necessity part, we only need to show that $\mathscr{V}(\Delta) \cap \mathscr{T}_1 = \{(1,1)\}$. So, assume that

$$\mathbf{w} = \Big[ \underbrace{\mathbf{w}_1}_{n-r} \Big| \mathbf{0} \Big]$$

is the GLPV of $\mathscr{M}$, and suppose that $\mathscr{V}(\Delta)$ intersects $\mathscr{T}_1$ at $(e^{i\omega_1}, e^{i\omega_2}) \neq (1,1)$. Then $n - r > 1$ and, by Theorem 4.1,

$$m_1 := \min_{1 \leqslant h \leqslant n-r} w_h, \qquad m_2 := \min_{1 \leqslant h \leqslant n-r} (1 - w_h)$$

are strictly positive quantities.

If we assume

$$\mathscr{X}_0' = \{\mathbf{x}'(h, -h) = \mathbf{w},\ h \in \mathbb{Z}\} \tag{4.18a}$$

as a sequence of initial probability vectors of $\mathscr{M}$, we obtain

$$\mathbf{x}'(h,k) = \mathbf{w} \tag{4.18b}$$

for any $(h,k)$ with $h + k \geqslant 0$. Consider now the following sequence of initial vectors:

$$\mathscr{X}_0'' = \mathscr{X}_0' + \frac{\mu}{2L}\mathscr{X}_0 \tag{4.20}$$

with $\mathscr{X}_0$ and $L$ defined in Lemma 4.6 and $\mu := \min(m_1, m_2)$.

The perturbation term $(\mu/2L)\mathscr{X}_0$ is small enough to guarantee that all vectors of the sequence $\mathscr{X}_0''$ are nonnegative. Moreover property (2) of Lemma 4.6 implies that the entries of each vector in $\mathscr{X}_0''$ sum to 1, so that $\mathscr{X}_0''$ may be considered as a sequence of probability vectors. The corresponding dynamical evolution of $\mathscr{M}$ is obtained as the superposition of (4.19), which provides a constant pattern in the half plane $\{(h,k): h + k \geqslant 0\}$, and (4.18), scaled down by $\mu/2L$, which does not converge to zero as $k + h \to 0$. This shows that $\mathbf{w}$ is not a LPV of $\mathscr{X}_0''$.

Conversely, suppose that $(1,1)$ is a regular point of $\mathscr{V}(\Delta)$ and $\mathscr{V}(\Delta) \cap \mathscr{T}_1 = \{(1,1)\}$. Then, by Lemma 4.3, $\lambda_0 = 1$ is a simple eigenvalue of the stochastic matrix $\mathbf{A} = a\mathbf{P} + (1 - a)\mathbf{Q}$. This implies that there exists a unique stochastic vector $\mathbf{w}$ satisfying (4.4), and hence a unique LPV. It remains to prove that $\mathbf{w}$ is a GLPV. To this purpose, all we need is to express the probability vectors as

$$\mathbf{x}(h,k) = \mathbf{w} + \mathbf{n}(h,k), \qquad \mathbf{n}(h,k) \in N$$

and to show that every initial sequence

$$\mathscr{N}_0 = \{\mathbf{n}(h, -h) \mid \mathbf{n}(h, -h) \in N,\ h \in \mathbb{Z}\} \tag{4.20}$$

evolves in the half plane $\{(h,k): h + k \geqslant 0\}$ so as to satisfy

$$\lim_{h + k \to +\infty} \mathbf{n}(h,k) = \mathbf{0}. \tag{4.21}$$

Indeed, an arbitrary vector in $N$ can be represented as a linear combination of the vectors $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n-1}$, introduced in (3.4):

$$\mathbf{n}(h,k) = \sum_{j=1}^{n-1} \hat{n}_j \mathbf{r}_j,$$

and the $(n-1)$-tuples $\hat{\mathbf{n}}(h,k) := [\hat{n}_1(h,k) \ \hat{n}_2(h,k) \ \cdots \ \hat{n}_{n-1}(h,k)]$ evolve according to an $(n-1)$-dimensional 2D system equation

$$\hat{\mathbf{n}}(h+1,k+1) = \hat{\mathbf{n}}(h,k+1)a\hat{\mathbf{P}}_{11} + \hat{\mathbf{n}}(h+1,k)(1-a)\hat{\mathbf{Q}}_{11}. \quad (4.22)$$

From the stability theory of 2D systems [7], we know that

$$\det\left[\mathbf{I}_{n-1} - az_1\hat{\mathbf{P}}_{11} - (1-a)z_2\hat{\mathbf{Q}}_{11}\right] \neq 0 \qquad \forall (z_1, z_2) \in \mathscr{P}_1 \quad (4.23)$$

is a necessary and sufficient condition guaranteeing $\hat{\mathbf{n}}(h,k) \to \mathbf{0}$ and, equivalently, $\mathbf{n}(h,k) \to \mathbf{0}$. Since in (3.7) the factor $1 - az_1 - (1-a)z_2$ vanishes at $(1,1)$, our hypotheses on $\mathscr{V}(\Delta)$ directly imply that the factor $\det[\mathbf{I}_{n-1} - az_1\hat{\mathbf{P}}_{11} - (1-a)z_2\hat{\mathbf{Q}}_{11}]$ is devoid of zeros in $\mathscr{P}_1$.

This completes the proof of the theorem. ∎

## REFERENCES


1 E. Fornasini and G. Marchesini, State space realization theory of two dimensional filters, *IEEE Trans. Automat. Control* AC-21(4):284–292 (1976).

2 E. Fornasini and G. Marchesini, Doubly indexed dynamical systems: State space models and structured properties, *Math. Systems Theory* 12:59–72 (1978).

3 S. Kung, B. Lévy, M. Morf, and T. Kailath, New results in 2D systems theory, parts I, II, *Proc. IEEE* 65(6):861–872, 945–961 (1977).

4 P. A. Devijver, Modeling of digital images using hidden Markov mesh random fields, in *Signal Processing: Theories and Applications* (J. L. Lacoune et al., Eds.), Elsevier Science, 1988, pp. 23–30.

5 A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic, 1979.

6 F. R. Gantmacher, *The Theory of Matrices*, Chelsea, 1959.

7 E. Fornasini and G. Marchesini, Stability analysis of 2D systems, *IEEE Trans. Circuits and Systems* CAS-27(12):1210–1217 (1980).

8 N. Jacobson, *Lectures in Abstract Algebra—Linear Algebra*, Van Nostrand, 1953.


*Received 7 June 1989; final manuscript accepted 12 December 1989*