A Sufficient Condition for a Unique Invariant Distribution of a Higher-Order Markov Chain

Bernhard C. Geiger

July 2, 2018

Abstract

We derive a sufficient condition for a $k$-th order homogeneous Markov chain $Z$ with finite alphabet $Z$ to have a unique invariant distribution on $Z^k$. Specifically, let $X$ be a first-order, stationary Markov chain with finite alphabet $X$ and a single recurrent class, let $g: X \to Z$ be non-injective, and define the (possibly non-Markovian) process $Y := g(X)$ (where $g$ is applied coordinate-wise). If $Z$ is the $k$-th order Markov approximation of $Y$, its invariant distribution is unique. We generalize this to non-Markovian processes $X$.

1 Introduction

We consider invariant distributions of a $k$-th order (i.e., “multiple”) Markov chain $Z := (Z_n)_{n \in \mathbb{N}_0}$ on a finite alphabet $Z$. For $k = 1$, the invariant distribution $\pi$ is a probability distribution on $Z$ and is unique if the Markov chain has a single recurrent class [1, Thm. 4.4.2]. If, in addition, the Markov chain is aperiodic, then the distribution of $Z_n$ converges to this invariant distribution as $n \to \infty$ [1, Thm. 4.3.7]. A fortiori, uniqueness and convergence are ensured if the Markov chain is regular, i.e., irreducible and aperiodic [2, Thm. 4.1.6].

For $k$-th order Markov chains, $k > 1$, two types of invariant distributions can be considered: A distribution $\pi$ on $Z$ and a distribution $\mu$ on $Z^k$.

The invariant distribution $\pi$ on $Z$ is related to the eigenvector problem of nonnegative tensors. Chang et al. showed that the eigenvector associated with the largest eigenvalue of an irreducible tensor is positive, but not necessarily simple [3, Thm. 1.4]. The results were used by Li and Ng in [4] to derive conditions under which there exists a unique distribution $\pi$ on $Z$ such that, for a $k$-th order Markov chain $Z$,

$$\forall z \in Z: \quad \pi_z = \sum_{z_0, \ldots, z_{k-1} \in Z} P(Z_k = z | Z_{k-1} = z_{k-1}, \ldots, Z_0 = z_0) \pi_{z_{k-1}} \cdots \pi_{z_0}. \quad (1)$$

Regarding convergence, Vladimirescu showed that if $Z$ is a regular second-order Markov chain, then there exists a distribution $\pi$ on $Z$ such that [5, eq. (7.1.3)]

$$\forall z, z_0, \ldots, z_{k-1} \in Z: \quad \pi_z = \lim_{n \to \infty} P(Z_n = z | Z_{k-1} = z_{k-1}, \ldots, Z_0 = z_0). \quad (2)$$

The more interesting case concerns invariant distributions $\mu$ on $Z^k$. Following Doob [6, p. 89], every $k$-th order Markov chain $Z$ on $Z$ can be converted
to a first-order Markov chain \( Z^{(k)} := (Z_n, \ldots, Z_{n+k-1})_{n \in \mathbb{N}_0} \) on \( Z^k \). Kalpazidou used this fact to characterize invariant distributions for Markov chains \( Z \) derived from weighted circuits [3, Prop. 7.2.2]. If \( Z^{(k)} \) is a regular first-order Markov chain, then \( Z \) is a regular \( k \)-th order Markov chain [3, Prop. 7.1.6]. Moreover, since in this case \( Z^{(k)} \) has a unique invariant distribution on \( Z^k \), so has \( Z \).

The converse, however, is not true: The regularity of \( Z^{(k)} \) does not follow from the regularity of \( Z \), and hence the uniqueness of an invariant distribution \( \mu \) on \( Z^k \) is not guaranteed even for a regular \( k \)-th order Markov chain \( Z \). To address this problem, Herkenrath discussed the uniform ergodicity of a second-order Markov process \( Z \) on a general alphabet \( Z \). Specifically, he defined the second-order Markov process \( Z \) to be uniformly ergodic if the first-order Markov process \( Z^{(2)} := (Z_{2n}, Z_{2n+1})_{n \in \mathbb{N}_0} \) on \( Z^2 \) is uniformly ergodic [2, Def. 4]. Herkenrath showed how sufficient and/or necessary conditions for uniform ergodicity (such as a strengthened Doeblin condition) carry over from \( Z \) to \( Z^{(2)} \) [2, Lem. 3]. He moreover presented sufficient conditions for uniform ergodicity of certain classes of second-order Markov processes, such as nonlinear autoregressive time series with absolutely continuous noise processes [7, Thms. 2-5]. Herkenrath moreover showed a relation between the invariant distributions \( \pi \) on \( Z \) and \( Z^2 \), respectively [7, Cor. 2]:

\[
\forall z \in Z: \quad \pi_z = \sum_{z' \in Z} \mu_{z,z'} = \sum_{z' \in Z} \mu_{z',z}. \tag{3}
\]

In this work, we present a sufficient condition for a \( k \)-th order homogeneous Markov chain \( Z \) on a finite alphabet \( Z \) to have a unique invariant distribution \( \mu \) on \( Z^k \). The condition is formulated via a function of a first-order Markov chain \( X \) with a single recurrent class. Since functions of Markov chains, so-called lumpings, usually do not possess the Markov property, one may need to approximate this lumping by a Markov chain with a given order. Assuming that this Markov approximation satisfies certain conditions, it can be shown that its invariant distribution is unique. We moreover generalize this result by letting \( X \) be a higher-order Markov chain and a non-Markovian process, respectively.

## 2 Problem Setting

We abbreviate vectors as \( z^k := (z_1, \ldots, z_k) \) and random vectors as \( Z^k := (Z_1, \ldots, Z_k) \). If the length of the vector is clear from the context, we omit indices. The probability that \( Z = z \) is written as \( p_Z(z) := P(Z = z) \); the conditional probability that \( Z_1 = z_1 \) given that \( Z_2 = z_2 \) is written as \( p_{Z_1|Z_2}(z_1|z_2) \). Stochastic processes are written as bold-faced letters, e.g., \( Z := (Z_n)_{n \in \mathbb{N}_0} \). We write sets with calligraphic letters. For example, the alphabet of \( Z \) is \( Z \). All processes and random variables are assumed to live on a finite alphabet, i.e., \( |Z| < \infty \). The complement of a set \( A \subseteq Z \) is \( Z^c := Z \setminus A \). Transition probability matrices are written in bold-face, too; whether a symbol is a matrix or a stochastic process will always be clear from the context. We naturally extend a function \( g: Z \rightarrow W \) from scalars to vectors by applying it coordinate-wise, i.e., \( g(z^k) := (g(z_1), \ldots, g(z_k)) \). Similarly, the preimage of a vector is the Cartesian product of the preimages, i.e., \( g^{-1}(w^k) := g^{-1}(w_1) \times \cdots \times g^{-1}(w_k) \).
A stochastic process $Z$ is a $k$-th order Markov chain with alphabet $Z$ if and only if

$$\forall n \geq k: \forall z^n_0 \in Z^{n+1}: \quad p_{Z_n | Z_0} (z_n | z_0^{n-1}) = p_{Z_{n-k} | Z_{n-k}^{n-1}} (z_n | z_{n-k}^{n-1}).$$

If the right-hand side of (4) does not depend on $n$, we can write

$$Q_{z_{n-k} \rightarrow z_n}^{Z} := p_{Z_0 | Z_0} (z_n | z_0^{n-1})$$

and call $Z$ homogeneous. We let $Q$ be a $|Z|^k \times |Z|$ matrix with entries $Q_{z_{n-k} \rightarrow z_n}$ and abbreviate $Z \sim \text{HMC}(k, Z, Q)$. Similarly, we define

$$Q_{z_{n-k} \rightarrow z_n}^{(n)} := p_{Z_{n-k} \rightarrow z_n} (z_n | z_0^{k-1})$$

and collect the corresponding values in the $|Z|^k \times |Z|$ matrix $Q^{(n)}$. Note that $Q^{(1)} = Q$.

We recall basic definitions for $k$-th order Markov chains $Z \sim \text{HMC}(k, Z, Q)$; the details can be found in, e.g., [5] Def. 7.1.1-7.1.4 or [1] Def. 4.2.2-4.2.7. A state $z \in Z$ is accessible from $z' \in Z$ (in short: $z' \rightarrow z$), if and only if

$$\forall u \in Z^{k-1}: \exists n = n(u, i, j): \quad Q_{z' \rightarrow z}^{(n)} > 0.$$  

If $z' \rightarrow z$ and $z \rightarrow z'$, then $z$ and $z'$ communicate (in short: $z \leftrightarrow z'$). A state $z$ is recurrent if $z \rightarrow z'$ implies $z' \rightarrow z$, otherwise $z$ is transient. The relation $\leftrightarrow$ partitions the set $\{z \in Z: \ z \leftrightarrow z\}$ into equivalence classes (called recurrent classes). The Markov chain $Z$ is irreducible if and only if $Z$ is the unique recurrent class; it is regular if and only if there exists $n \geq 1$ such that $Q^{(n)}$ is a positive matrix.

A $k$-th order Markov chain $Z \sim \text{HMC}(k, Z, Q)$ can be converted to a first-order Markov chain on $Z^k$. Let $Z^{(k)} := (Z^k_u)_{u \in \mathbb{N}_0}$. Then, $Z^{(k)} \sim \text{HMC}(1, Z^k, P)$, where [1] p. 89

$$\forall z_0^{k-1}, z_0^k \in Z^k: \quad P_{z_0^{k-1} \rightarrow z_0^k} = \begin{cases} Q_{z_0^{k-1} \rightarrow z_0^k}, & z_0^k = z_0, \ldots, z_{k-1}^k = z_{k-2} \\ 0, & \text{else.} \end{cases}$$

(8)

**Definition 1** (Invariant Distribution). Let $Z \sim \text{HMC}(k, Z, Q)$. A distribution $\mu$ on $Z^k$ is invariant if and only if

$$\forall z_1^k \in Z^k: \quad \mu_{z_1^k} = \sum_{z_0 \in Z} \mu_{z_0^{k-1}} Q_{z_0^{k-1} \rightarrow z_0^k}.$$  

(9)

It can be shown that a distribution $\mu$ on $Z^k$ is invariant for $Z$ if and only if it is invariant for $Z^{(k)}$. If $Z^{(k)}$ has a single recurrent class, then this $\mu$ is unique [1] Thm. 4.4.2]. The following example illustrates that even if $Z$ is regular, $Z^{(k)}$ may have multiple recurrent classes:
Example 1. Let $Z \sim \text{HMC}(2, \{1, 2, 3, 4\}, Q)$, where

$$Q = \begin{pmatrix}
1 & 2 & 3 & 4 \\
1,1 & 0.5 & 0.5 & 0 \\
1,2 & 0 & 0 & 1 & 0 \\
1,3 & 0 & 1 & 0 & 0 \\
1,4 & 0 & 0 & 1 & 0 \\
2,1 & 0 & 0 & 0.5 & 0.5 \\
2,2 & 0 & 0.5 & 0.5 & 0 \\
2,3 & 0.5 & 0 & 0 & 0.5 \\
2,4 & 1 & 0 & 0 & 0 \\
3,1 & 0 & 1 & 0 & 0 \\
3,2 & 1 & 0 & 0 & 0 \\
3,3 & 0 & 0.5 & 0.5 & 0 \\
3,4 & 1 & 0 & 0 & 0 \\
4,1 & 0 & 1 & 0 & 0 \\
4,2 & 0 & 1 & 0 & 0 \\
4,3 & 0 & 1 & 0 & 0 \\
4,4 & 0 & 0 & 0.5 & 0.5
\end{pmatrix}. \quad (10)$$

This Markov chain $Z$ is regular since $Q^{(10)} > 0$. $Z$ is such that, depending on the initial states, we either observe sequences $1 - 2 - 3 - 4 - 1$ and $1 - 2 - 3 - 1$ or sequences $1 - 4 - 3 - 2 - 1$ and $1 - 3 - 2 - 1$. It follows that $Z^{(2)}$ has transient states $\{(1, 1), (2, 2), (2, 4), (3, 3), (4, 2), (4, 4)\}$ and two recurrent classes:

$\{(1, 2), (2, 3), (3, 1), (3, 4), (4, 1)\}$

and

$\{(1, 3), (3, 2), (2, 1), (1, 4), (4, 3)\}$.

Hence, there is no unique invariant distribution $\mu$ satisfying $\mathcal{M}$.

We define the Markov approximation of a non-Markovian process:

Definition 2 (k-th order Markov Approximation). Let $W$ be a stationary stochastic process on the finite alphabet $Z$. The k-th order Markov approximation of $W$ is a k-th order Markov chain $Z \sim \text{HMC}(k, Z, Q)$ with transition matrix $Q$ with entries

$$\forall z_0^k \in Z^{k+1}: \quad Q_{z_0^{k-1} \rightarrow z_k} = \begin{cases} \frac{p_{W_k}(z_k)}{p_{W_0^{k-1}}(z_0^{k-1})}, & p_{W_0^{k-1}}(z_0^{k-1}) > 0 \\ \frac{1}{2^n}, & \text{else}. \end{cases} \quad (11)$$

The approximation $(11)$ can be justified via information theory. We define the relative entropy rate between $W$ and a k-th order Markov chain $M \sim \text{HMC}(k, Z, \bar{Q})$ as [8] p. 266

$$\mathbb{D}(W \| M) := \lim_{n \to \infty} \frac{1}{n} \sum_{z_0^{n-1} : p_{W_0^{n-1}}(z_0^{n-1}) > 0} p_{W_0^{n-1}}(z_0^{n-1}) \log \frac{p_{W_0^{n-1}}(z_0^{n-1})}{p_{M_0^{n-1}}(z_0^{n-1})}. \quad (12)$$

whenever the limit exists and is finite. Then, one can show that [3] Cor. 10.4

$$\inf_{\mathcal{Q}} \mathbb{D}(W \| M) = \mathbb{D}(W \| Z) \quad (13)$$
where $Z \sim \text{HMC}(k, Z, Q)$ with $Q$ defined in (11). Note that the minimizer of (13) is not unique: Those sequences $z_0^{n-1}$ that contain a subsequence $z_{t}^{\ell+k-1}$ with $p_{W_{0}}^{k-1}(z_{t}^{\ell+k-1}) = 0$ are not part of the sum in (12), hence the choice of $Q_{z_{t}^{\ell+k-1}}$ is immaterial. We will justify our particular choice later. The following example shows that approximating $W$ may lead to counterintuitive results if $W$ is a first-order Markov chain:

Example 2. Let $W \sim \text{HMC}(1, \{1, 2, 3\}, P)$ with $P_{1 \rightarrow 3} = P_{2 \rightarrow 1} = P_{3 \rightarrow 1} = 0$ and all other entries $P_{x_{0} \rightarrow x_{1}} > 0$. $W$ has a single recurrent class $\{2, 3\}$ and a transient state $\{1\}$, hence the unique invariant distribution $\pi$ satisfies $\pi_{1} = 0$. Let $Z \sim \text{HMC}(1, \{1, 2, 3\}, Q)$ be the first-order Markov approximation of $W$. Since $\pi_{1} = 0$, we have that

$$\exists x \in \{1, 2, 3\}: \quad Q_{1 \rightarrow x} = \frac{1}{3} \neq P_{1 \rightarrow x}$$

and $Z \neq W$. Nevertheless, we get $\mathbb{D}(W\|Z) = 0$.

Moreover, the $k$-th order Markov approximation of $W$ may even fail to be Markov of any order smaller than $k$. Indeed, we may have that

$$\exists x^{k} \in \{2, 3\}^{k}: \quad Q_{(1, x_{1}^{k-1}) \rightarrow x_{k}} = \frac{1}{3} \neq Q_{(2, x_{1}^{k-1}) \rightarrow x_{k}} = P_{x_{k-1} \rightarrow x_{k}}$$

Lemma 1. Let $W$ be a stationary stochastic process on the finite alphabet $Z$ and let $Z \sim \text{HMC}(k, Z, Q)$ be its $k$-th order Markov approximation. Then, $p_{W_{0}}^{k-1}$ is an invariant distribution of $Z$.

Proof. If $p_{W_{0}}^{k-1}$ is invariant, then for every $z_{0}^{k} \in Z^{k}$,

$$p_{W_{0}}^{k-1}(z_{0}^{k}) = \sum_{z_{0} \in Z} p_{W_{0}}^{k-1}(z_{0}^{k-1})Q_{z_{0}^{k-1} \rightarrow z_{k}}$$

$$= \sum_{z_{0} \in Z} p_{W_{0}}^{k}(z_{0}) = p_{W_{0}}^{k}(z_{1}^{k}) = p_{W_{0}}^{k-1}(z_{1}^{k})$$

where (a) is because those $z_{0}$ for which $p_{W_{0}}^{k-1}(z_{0}^{k-1}) = 0$ do not influence the sum, and where (b) is because $W$ is stationary.

3 Main Result

We present our main result:

Theorem 1. Let $X \sim \text{HMC}(1, X, P)$ have a single recurrent class and invariant distribution $\pi$. Let $X$ be stationary, i.e., $p_{X_{0}} = \pi$. Let $g: X \rightarrow Y$ where $1 \leq |Y| \leq |X|$, i.e., $g$ may be non-injective. Define $Y$ via $Y_{n} = g(X_{n})$, and let $Z \sim \text{HMC}(k, Y, Q)$ be the $k$-th order Markov approximation of $Y$ with $Q$ given in (11). Then, $Z$ has a unique invariant distribution $\mu$ on $Y^{k}$ satisfying

$$\forall y_{0}^{k-1} \in Y: \quad \mu_{y_{0}^{k-1}} = p_{y_{0}^{k-1}}(y_{0}^{k-1}) = \sum_{x_{0} \in X^{k}} \pi_{x_{0}} \prod_{\ell=1}^{k} P_{x_{\ell-1} \rightarrow x_{\ell}}$$

$$\exists y_{0}^{k-1} \in Y: \quad \mu_{y_{0}^{k-1}} = p_{y_{0}^{k-1}}(y_{0}^{k-1}) = \sum_{x_{0} \in X^{k}} \pi_{x_{0}} \prod_{\ell=1}^{k} P_{x_{\ell-1} \rightarrow x_{\ell}}$$

(17)
Theorem 1 holds also for bijective functions. For these functions, however, \( Y \) is a first-order Markov chain and we may face counterintuitive results as illustrated in Example 2. A similar situation may occur if \( X \) is lumpable w.r.t. the non-injective function \( g \), i.e., if \( Y \) is a Markov chain \([2] \S 6.3\). Hence, Theorem 1 is useful mainly in situations where \( Y \) is not Markov of any order.

Theorem 1 holds for the definition of \( Q \) in (11) where the conditional distribution for a conditioning event with zero probability is chosen as the uniform distribution. More generally, it holds if the uniform distribution is replaced by any positive probability vector on \( Z \). This positivity constraint cannot be dropped, however, as the following example illustrates:

**Example 3.** Let \( X \sim \text{HMC}(1, \{1, 2, 3\}, P) \) with \( P_{1 \rightarrow 1} = 0 \) and all other entries \( P_{x_0 \rightarrow x_1} > 0 \). Let \( g: \{1, 2, 3\} \rightarrow \{1, 2\} \) be such that \( g(1) = 1 \) and \( g(2) = g(3) = 2 \). Suppose we want to model \( Y \) by a second-order Markov chain \( Z \). We get

\[
Q = \begin{pmatrix}
1 & 2 \\
1, 1 & ? \\
1, 2 & q & 1 - q \\
2, 1 & 0 & 1 \\
2, 2 & p & 1 - p
\end{pmatrix}
\]

(18)

where \( p, q \in (0, 1) \) and where the ? indicates that the distribution of \( Y \) does not tell us how to choose these probabilities (the event \( Y_0^1 = (1, 1) \) occurs with probability zero).

Choosing \( Q_{(1,1)\rightarrow 1} = 1 \) (and hence \( Q_{(1,1)\rightarrow 2} = 0 \)) results in \( Z^{(2)} \) having the two recurrent classes \( \{(1, 1)\} \) and \( \{(1, 2), (2, 1), (2, 2)\} \). The invariant distribution is thus not unique. Indeed, for every invariant distribution, the state \((1, 1)\) has the same probability as it has in the initial distribution.

The following result extends Theorem 1 to the scenario where \( Y \) is the function of a \( k \)-th order Markov chain.

**Corollary 1.** Let \( X \sim \text{HMC}(\ell, \mathcal{X}, P) \) be such that \( X^{(\ell)} \) has a single recurrent class and the unique invariant distribution \( \pi \). Let \( X^{(\ell)} \) be stationary, i.e., \( p_{X^{(\ell-1)}} = \pi \). Let \( g: \mathcal{X} \rightarrow \mathcal{Y} \). Define \( Y \) via \( Y_n = g(X_n) \), and let \( Z \sim \text{HMC}(k, \mathcal{Y}^1, Q) \) be the \( k \)-th order Markov approximation of \( Y \) with \( Q \) given in (11). Then, \( Z \) has a unique invariant distribution \( \mu \) on \( \mathcal{Y}^k \).

**Proof.** Note that \( X^{(\ell)} \) is a first-order Markov chain on \( \mathcal{X}^\ell \). By assumption, it has a single recurrent class. We define the function \( g^{(\ell)}: \mathcal{X}^{\ell} \rightarrow \mathcal{Y} \) as

\[
\forall x_0^{\ell-1} \in \mathcal{X}^{\ell}: \quad g^{(\ell)}(x_0^{\ell-1}) = g(x_0).
\]

(19)

Hence, \( Y_n = g(X_n) = g^{(\ell)}(X_n^{n+\ell-1}) \). Theorem 1 completes the proof. \( \square \)

The condition in Corollary 1 holds for every pair of integers \( \ell \) and \( k \). However, requiring that \( X \) has a unique invariant distribution on \( \mathcal{X}^{\ell} \) is too restrictive if \( k < \ell \). The next result shows that it suffices if the \( k \)-th order Markov approximation of \( X \) has a unique invariant distribution on \( \mathcal{X}^k \) to ensure that \( Z \) has a unique invariant distribution on \( \mathcal{Y}^k \). We can thus drop the condition that \( X \) is Markov by showing that the order of Markov approximation and projection through a non-injective function commutes in the sense of Figure 1.
Figure 1: The order of Markov approximation and projection through a non-injective function commutes. Dashed arrows labeled with “k-App.” denote the k-th order Markov approximation in the sense of Definition 2, solid arrows labeled with “g” denote a projection of the stationary process through the non-injective function g. While the (generally non-Markovian) processes Y and ˜Y differ, their k-th order Markov approximations are identical.

Corollary 2. Let X be a stationary stochastic process on the finite alphabet X. Let g : X → Y. Define Y via Y_n = g(X_n), and let Z ~ HMC(k, Y, Q) be the k-th order Markov model of Y with Q given in (11). If the k-th order Markov approximation M of X is such that M(k) has a single recurrent class, then Z has a unique invariant distribution µ on Y^k.

Proof. Let M ~ HMC(k, X, P) be the k-th order Markov approximation of X. If M(k) has a single recurrent class, its invariant distribution is unique and equals, by Lemma 1, \pi = p_{X_{k-1}}. We set M^*_k = \pi such that M is stationary and define ˜Y via ˜Y_n = g(M_n). If we can show that the k-th order Markov approximation of ˜Y coincides with the k-th order Markov approximation of Y, then the proof follows from Corollary 1. It therefore suffices to show that p_{Y^k} = p_{˜Y^k}. We have

\[\forall y^k_0 \in \mathcal{Y}^{k+1}, \quad p_{\tilde{Y}^{k-1}}(y^k_0) = \sum_{x^k_0 \in g^{-1}(y^k_0)} p_{M^*_k}(x^k_0)\]

\[= \sum_{x^k_0 \in g^{-1}(y^k_0)} p_{X^{k-1}}(x^k_0) p_{x^{k-1}} \rightarrow x_k \]  \hspace{1cm} (20)

\[= \sum_{x^k_0 \in g^{-1}(y^k_0)} p_{X^{k-1}}(x^k_0) \frac{p_{x^k}(x^k_0)}{p_{x^{k-1}}(x^{k-1}_0)} \]  \hspace{1cm} (21)

\[= \sum_{x^k_0 \in g^{-1}(y^k_0)} p_{X^{k-1}}(x^k_0) = p_{Y^k}(y^k_0) \]  \hspace{1cm} (22)

where in (a) we used the fact that p_{X^{k-1}} is the unique invariant (thus stationary) distribution of M and where (b) follows because those \[x^k_0 \in g^{-1}(y^k_0)\] for which \[p_{X^{k-1}}(x^{k-1}_0) = 0\] do not influence the sum. This completes the proof.  \[\Box\]
4 Proof of Theorem

That \( \mu = p_{Y_{n-1}} \) is an invariant distribution for \( Z \) follows from Lemma 1. To show that this invariant distribution is unique, we show that the first-order Markov chain \( Z(k) \) has the single recurrent class

\[
S := \{ y_0^{k-1} \in Y : p_{Y_0^{k-1}}(y_0^{k-1}) > 0 \}.
\]

(24)

We show in Section 4.2 that all states in \( S \) communicate by showing that for all \( y, y' \in S \) we have \( y \to y' \) (hence, \( y \leftrightarrow y' \)). We show in Section 4.3 that \( S \) is a recurrent class by showing that, for \( y \in S \) and \( y' \in S' \), we have \( y \not\sim y' \). Finally, in Section 4.4 we show that the states in \( S' \) are transient by showing that for every \( y' \in S' \) there is an \( y \in S \) such that we have \( y' \to y \). This completes the proof that \( S \) is the single recurrent class of \( Z(k) \), from which the uniqueness of \( \mu \) follows.

4.1 All states in \( S \) communicate

Lemma 2. For every \( n \geq 0 \) and every \( y_0^{k-1} \in Y^k \) such that \( p_{Y_0^{k-1}}(y_0^{k-1}) > 0 \), we have

\[
\forall y_0^{k+n} \in Y^{k+n+1}: \quad p_{Z^{k+n}|Z_0^{k-1}}(y_0^{k+n}|y_0^{k-1}) = 0 \Rightarrow p_{Y_0^{k+n}|Y_0^{k-1}}(y_0^{k+n}|y_0^{k-1}) = 0.
\]

(25)

Proof. This result is an immediate consequence of [8, Thm. 10.1, eq. (10.8)]. There, it was shown that with \([11]\) and with setting \( p_{Z^{k-1}|y_0^{k-1}} = p_{Y^{k-1}|y_0^{k-1}} \), one gets for every \( n \geq 0 \)

\[
\forall y_0^{k+n} \in Y^{k+n+1}: \quad p_{Z^{k+n}|Z_0^{k-1}}(y_0^{k+n}|y_0^{k-1}) = 0 \Rightarrow p_{Y_0^{k+n}|Y_0^{k-1}}(y_0^{k+n}|y_0^{k-1}) = 0.
\]

(26)

The proof follows by dividing both sides by \( p_{Y_0^{k-1}}(y_0^{k-1}) > 0 \).

We show that all states in \( S \) communicate by showing that, for every pair \( y, y' \in S \), we have \( y \to y' \). With Lemma 2 it thus suffices to show that for every pair \( y, y' \in S \) there exists an \( n = n(y, y') > 0 \) such that

\[
p_{Y_{n+k-1}|Y_{k-1}}(y'|y) > 0.
\]

(27)

We can write this as

\[
p_{Y_{n+k-1}|Y_{k-1}}(y'|y) = \sum_{x_0^{k-1} \in g^{-1}(y)} p_{X_n^{k-1}|X_0^{k-1}}(x_n^{k-1}|x_0^{k-1}) p_{X_0^{k-1}|Y_0^{k-1}}(x_0^{k-1}|y).
\]

\[
= \sum_{x_0^{k-1} \in g^{-1}(y)} p_{X_{n+k-1}|X_{k-1}}(x_{n+k-1}|x_{n+k-1}) p_{X_0^{k-1}|Y_0^{k-1}}(x_0^{k-1}|y).
\]

(28)

Since \( y \in S \) and \( y' \in S \), there exist \( x_0^{k-1} \in g^{-1}(y) \) and \( x_n^{n+k-1} \in g^{-1}(y') \) such that \( p_{X_0^{k-1}}(x_0^{k-1}) > 0 \) and \( p_{X_0^{k-1}}(x_0^{k-1}) > 0 \). It follows that \( x_{k-1} \) and
are elements of the single recurrent class of $X$. Moreover, it follows that $p_{X_n^k|X_0^k}(x_0^{k-1}|y) > 0$ and that

$$p_{X_n^k}(x_0^{n+k-1}) = \pi_{x_0^n} \prod_{\ell=n+1}^{n+k-1} P_{x_\ell^k \rightarrow x_\ell^k} > 0. \tag{29}$$

Since $X$ is Markov, the first term in the sum in (28) can be written as

$$\frac{p_{X_n^k}(x_0^{n+k-1})}{\prod_{\ell=\max\{k,n\} + 1}^{n+k-1} P_{x_\ell^k \rightarrow x_\ell^k}} \left( P_{x_\ell^k \rightarrow x_\ell^k} > 0 \right). \tag{30}$$

Since $x_{k-1}$ and $x_{\max\{k,n\}}$ are in the same recurrent class, it follows that there exists an $n = n(x_{k-1}, x_{\max\{k,n\}})$ such that the first term in the product is positive; the second term is positive by (29). Combined with $p_{X_n^k|Y_0^{k-1}}(x_0^{k-1}|y) > 0$ it follows that at least one summand in (28) is positive, from which (27) follows.

4.2 $S$ is a recurrent class

We already showed that, for all $y, y' \in S$, $y \leftrightarrow y'$. To show that $S$ is a recurrent class, i.e., an equivalence class under the relation “$\leftrightarrow$”, we must show that for $y \in S$ and $y' \in S^c$, we have $y \not\leftrightarrow y'$, i.e.,

$$\forall y \in S, y' \in S^c: \forall n \geq 1: p_{X_n^k|Z_0^{k-1}}(y'|y) = 0. \tag{31}$$

Suppose the contrary is true, i.e., there exists a $y_0^{k-1} \in S$ and an $n \geq 1$, such that for a $y_0^{n+k-1} \in S^c$ we have $p_{X_n^{n+k-1}|Z_0^{k-1}}(y_0^{n+k-1}|y_0^{k-1}) > 0$. From this follows that there exists at least one sequence $y_k^{n+k-1} \in Y^n$ such that

$$p_{X_n^{n+k-1}|Z_0^{k-1}}(y_k^{n+k-1}|y_0^{k-1}) = \prod_{\ell=k}^{n+k-1} Q_{y_\ell^{k-1} \rightarrow y_\ell} > 0. \tag{32}$$

Since, by assumption, $y_0^{k-1} \in S$ and $y_0^{n+k-1} \in S^c$, there must be an $\ell \in \{k, \ldots, k + n - 1\}$ such that $y_\ell^{k-1} \in S$ and $y_\ell^{k+1} \in S^c$. But

$$p_{Y_0^{k-1}}(y_\ell) \overset{(a)}{=} p_{Y_0^{k-1}}(y_\ell^\ell) = \sum_{y_\ell^\ell \in Y_\ell^\ell} p_{Y_0^{k-1}}(y_\ell^\ell) \overset{(b)}{=} p_{Y_0^{k-1}}(y_\ell^{k-1}) = p_{Y_0^{k-1}}(y_\ell^{k-1}) p_{Y_0^{k-1}}(y_\ell^{k-1}) \overset{(c)}{=} Q_{y_\ell^{k-1} \rightarrow y_\ell} p_{Y_0^{k-1}}(y_\ell^{k-1}) \overset{(d)}{=} 0. \tag{33}$$
where (a) is due to stationarity of \( Y \) and where (b) is because \( y_{\ell-1}^{k-1} \in S \) and because of (11) in Definition 2. Finally, (c) follows from \( y_{\ell-1}^{k-1} \in S \) and (32). Hence \( y_{k+1}^\ell \in S \), a contradiction.

4.3 All states in \( S^c \) are transient

To show that all states in \( S^c \) are transient, we must show that for every \( y \in S^c \) there is at least one \( y' \in S \) such that \( y \rightarrow y' \). Since we showed in the last section that \( y' \not\rightarrow y \), it follows that \( y \) is transient.

Suppose \( y_{k-1}^0 \in S^c \). From (11) we get

\[
\forall k \in Y: \quad p_{Z_k|Z_{k-1}}^{y_0^{k-1}}(y_k|y_0^{k-1}) = \frac{1}{|Y|}.
\] (33)

If for at least one \( y_k \) we have \( y_k^1 \in S \), then we are done. If for every \( y_k \) we have \( y_k^1 \in S^c \), then

\[
\forall k \in Y: \quad p_{Z_k|Z_{k-1}}^{y_0^{k-1}}(y_k^{k+1}|y_0^{k-1}) = \frac{1}{|Y|^2}.
\] (34)

If for at least one \( y_k^{k+1} \) we have \( y_2^{k+1} \in S \), then we are done. Otherwise, we repeat above procedure. Eventually, if for every \( y_k^{2k-2} \in Y^{k-1} \) we have \( y_{k-2}^{2k-2} \in S^c \), then

\[
\forall k \in Y: \quad p_{Z_k|Z_{k-1}}^{y_0^{k-1}}(y_k^{2k-1}|y_0^{k-1}) = \frac{1}{|Y|^k}.
\] (35)

But at least one \( y_k^{2k-1} \in Y^k \) must be such that \( y_{k-1}^{2k-1} \in S \), since \( S \) is non-empty. Thus, from every \( y \in S^c \) there is at least one \( y' \in S \) such that \( y \rightarrow y' \).

5 Conclusion and Outlook

A \( k \)-th order homogeneous Markov chain \( Z \) with finite alphabet \( Z \) has a unique invariant distribution on \( Z^k \) if the first-order Markov chain \( Z(k) \) has a single recurrent class. We presented a sufficient condition for this to be the case: \( Z \) has a unique invariant distribution on \( Z^k \) if it is the \( k \)-th order Markov approximation of a function of a first-order Markov chain \( X \) with a single recurrent class. This condition has practical relevance in, e.g., state space reduction for Markov chains, e.g., [9, 10]. We generalized our result to \( X \) being a Markov chain of any order and to \( X \) being not Markov at all.

Example 2 suggests that our \( k \)-th order Markov approximation in Definition 2 leads to counterintuitive results if the process to be approximated is a Markov chain of order smaller than or equal to \( k \). Future work shall investigate whether a different choice of \( Q \) in Definition 2 can alleviate this problem while still ensuring that Theorem 1 holds.

References

References

[1] R. G. Gallager, Stochastic Processes. Theory for Applications, Cambridge University Press, Cambridge, 2013.
[2] J. G. Kemeny, J. L. Snell, Finite Markov Chains, 2nd Edition, Springer, 1976.

[3] K. C. Chang, K. Pearson, T. Zhang, Perron-Frobenius theorem for non-negative tensors, Commun. Math. Sci. 6 (2) (2008) 507–520.

[4] W. Li, M. K. Ng, On the limiting probability distribution of a transition probability tensor, Linear and Multilinear Algebra 62 (3) (2014) 362–385.

[5] S. L. Kalpazidou, Cycle Representations of Markov Processes, 2nd Edition, Vol. 28 of Stochastic Modelling and Applied Probability, Springer, New York, 2006.

[6] J. L. Doob, Stochastic Processes, Wiley Classics Library, Wiley-Interscience, New York, NY, 1990.

[7] U. Herkenrath, On the uniform ergodicity of Markov processes of order 2, J. Appl. Prob 40 (2) (2003) 455–472.

[8] R. M. Gray, Entropy and Information Theory, Springer, New York, NY, 1990.

[9] B. C. Geiger, Y. Wu, Higher-order optimal Kullback-Leibler aggregation of Markov chains, in: Proc. Int. ITG Conf. on Systems, Communications and Coding (SCC), Hamburg, 2017, pp. 1–6, open-access: arXiv:1608.04637 [cs.IT].

[10] K. Deng, P. G. Mehta, S. P. Meyn, Optimal Kullback-Leibler aggregation via spectral theory of Markov chains 56 (12) (2011) 2793–2808.

**Funding**

The work was funded by the Erwin Schrödinger Fellowship J 3765 of the Austrian Science Fund.