Spectral Analysis of Uncertain Dynamical Systems for Pattern Classification and Anomaly Detection

Najah F Ghalyan\textsuperscript{a}, Asok Ray\textsuperscript{a,b,*}

\textsuperscript{a}Department of Mechanical Engineering, The Pennsylvania State University, University Park, PA 16802, USA
\textsuperscript{b}Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA

Abstract

Spectral analysis has been widely used as a statistical tool for signal analysis and pattern recognition in uncertain dynamical systems, where the underlying process is often characterized by eigenfunctions of relevant stochastic transformations. Such a characterization is useful for various aspects of statistical data analysis (e.g., dimensionality reduction, data compression, and feature extraction). In this context, the notion of ergodicity has a strong influence on the spectral structure of semigroups of measure-preserving transformations. This paper investigates the spectral properties of ergodicity for applications to signal analysis, pattern recognition, and anomaly detection, where a measure-preserving transformation (MPT)-based methodology is proposed for construction of non-stationary probabilistic finite state automata (PFSA) to model the underlying stochastic process by using short-length time series of measurements. The resulting PFSA models are non-homogeneous Markov, in which the spectral properties of the MPT depend on several parameters that include window length of the observed time series for constructing the PFSA. The paper also develops an MPT-based metric to quantify the divergence of the evolving PFSA from the nominal PFSA; this information is then used for pattern classification.

*Corresponding author: Tel.: +1-814-865-6377;
Email address: axr2@psu.edu (Asok Ray)

Preprint submitted to Physica A: Statistical Mechanics and its Applications October 6, 2019
and anomaly detection with low-delay tolerance. The MPT-based methodology has been validated with experimental data generated from a laboratory apparatus that deals with detection of thermoacoustic instabilities in combustion processes, which are known to have chaotic characteristics on the time-scale of milliseconds. In this application, the concept of MPT has been used to develop a novel symbolic time series analysis (STSA)-based detection method, whose performance is evaluated by comparison with a standard STSA-based method on the same experimental data. The results consistently show superior performance of the proposed MPT-based STSA.

Keywords: Symbolic time series analysis; Spectral analysis; Ergodicity.

1. Introduction

Detection of anomalies and behavioral changes is important in diverse physical applications from the perspectives of rigorous statistical analysis; examples of such physical applications are radar systems, condition-based maintenance, durability of mechanical structures, and malware detection (e.g., and references therein). A general framework for anomaly detection and change point detection involves selection of an appropriate cost functional that measures goodness-of-fit of the generated time series to a (nominal) model, and identifying signal segments for which the goodness of fit satisfies a threshold criterion. In this context, Bai and Perron addressed the problem of estimating the change points by introducing a computationally-efficient algorithm that finds the global minimum of the sum of squared residuals using dynamic programming, where the technical approach considers the entire time series for detection of change points and thus imposes no constraint on the amount of delay in making these decisions. In many physical applications (e.g., thermo-fluid and combustion processes), changes need to be detected by using short observation windows in order to reduce the encountered delay of detection.

The research topic of low-delay detection has attracted the attention of many researchers over the last six decades. One of the standard methods, in this re-
gard, is the cumulative sum (CUSUM) technique, developed by Page [6], which
has been widely used for detection of change points [7, 8]. In the same con-
text, Shiryaev [9] established a Bayesian method to address the problem of fast
sequential detection of change (disorder) in time series using the optimal stop-
pling theory. The aim of Shiryaev’s work was to achieve change point detection
with minimum average delay to detection and maximum expectation of inter-
vals between consecutive false alarm signals. However, this work was limited
to identical and independent distribution (iid) of the observed time series; this
restriction has been subsequently addressed by Tartakovsky and Veeravalli [10]
who established a general asymptotic change-point detection theory that is not
limited to iid observations.

The above discussion largely applies to cases where the change in the time
series occurs as a result of abrupt variation in the underlying model structure.
Nevertheless, in many applications, changes in the time series may happen due
to gradual degradation in the underlying dynamical system. Detection of such
gradual changes is more challenging than those due to abrupt changes. Tools
of hidden Markov modeling have been widely used for detection of both abrupt
changes and gradually evolving anomalies in diverse applications (e.g., speech
recognition [11], electronic systems [12], bioinformatics [13, 14], target detection
and tracking [15], facial recognition [16], and brain imaging [17]). In this set-
ting, the (nominal) hidden Markov model (HMM) is trained by using observed
time series that is known to represent the nominal behavior of the uncertain
dynamical system under consideration. If a change occurs in the time series,
then the likelihood of the new observed subsequence under the nominal HMM
is expected to deviate significantly from the typical likelihood [18]. Relying on
a properly chosen threshold [4], which could be fixed (e.g., based on a specified
false detection rate), one can decide whether a change has occurred in the time
series or not.

In a similar context, the concept of symbolic time series analysis (STSA)
has been used by many researchers (e.g., [19, 20, 21]) for constructing Markov-
chain models from the respective observed time series. In this framework, a
(finite-length) time series is partitioned for conversion into a string of symbols from a (finite-cardinality) alphabet $\mathcal{A}$ (e.g., $[22, 23, 24, 25]$). Subsequently, a probabilistic finite state automaton (PFSA) is constructed from the symbol string (e.g., $[26, 27, 28]$), in which the probability distribution of the emitted symbols depends upon the immediately preceding $D$ symbols, where $D$ is a positive integer called Markov depth. Such a PFSA is called a $D$-Markov machine, which has found diverse applications in pattern classification and anomaly detection (e.g., $[28, 22, 21]$). The main distinction between HMMs and $D$-Markov machines is that the algebraic structure of state transition in HMMs is non-deterministic while it is restricted to be deterministic in $D$-Markov machines $[27, 28, 21]$. This restriction yields significant computational advantages of $D$-Markov machines over HMMs $[26, 27, 28]$. Moreover, training HMMs is typically done statistically by using an expectation maximization tool (e.g., Baum-Welch algorithm $[29]$). In addition to its iterative computation cost, such a method only guarantees convergence to a point, having zero gradient with respect to the HMM parameters, and therefore this point could be a poor local optimum. In contrast, the deterministic algebraic structure of $D$-Markov machines makes the modeling process much simpler and less prone to the local optimum issue, where the model can be trained using the frequency counting method $[28]$, for example.

In the above setting, selection of the window length, $L$, of the time series used to construct the PFSA in $D$-Markov machines largely depends on the Markov depth $D$, the alphabet size $|\mathcal{A}|$, and the nature of the particular underlying process generating the time series $[28, 30]$. To find a lower bound on the window length of the time series required to estimate the PFSA parameters, one may consider an increasing sequence of window size. Under the assumption of statistical stationarity $[30, 31]$, the stochastic (state transition probability) matrix converges to a constant matrix; as a consequence, the window length may become arbitrarily large $[21]$. Thus, one may choose the smallest window length at which the stochastic matrix tends to be approximately time-invariant. The resulting model in this case would be a time-homogeneous Markov chain $[32]$. 
However, this scenario would typically require a large value of $L$, which could be infeasible in many physical applications where decisions need to be made with low-delay tolerance. In this context, an application example is presented below.

The application example, investigated in this paper, focuses on monitoring and (near-real-time) active control of thermoacoustic instabilities (TAI) in combustion processes, which are usually caused by spontaneous excitation of one or more natural modes of the acoustic waves. The TAI phenomena are typically manifested by large-amplitude self-sustained (possibly chaotic) pressure oscillations in the combustion chamber, which may lead to damage in mechanical structures if the pressure oscillations match one of the natural frequencies of the combustor. The time scales of TAI are on the order of milliseconds, which must be mitigated by fast actuation of control signals.

In the analysis of uncertain dynamical systems, ergodicity of the underlying stochastic process is crucial for generation of the corresponding symbolic sequence. In this context, Rohlin theorem states that a dynamical system can be uniquely described by its symbolic representation if the system is ergodic. Moreover, the notion of ergodicity is very conducive for analyses of time series data (e.g., coding and signal processing that operate sequentially on disjoint blocks of the data stream). An ergodic process serves as a source of information that, if observed for a sufficiently long time, will emit strings of symbols, which are accurate representations of true signals in the sense of having its computed entropy close to the true entropy of the source; fortunately, ergodicity arises naturally in many physical processes. In fact, the idea of information sources possessing the ergodic property is so natural that one may have difficulty in identifying sources that are not ergodic.

Earlier publications have reported identification of two aforementioned parameters of $D$-Markov machines, namely, Markov depth $D$ and the alphabet size $|\mathcal{A}|$ for construction of PFSA. This paper proposes a methodology for identifying the remaining aforesaid parameter, namely, the minimum window length $L$, for symbolic analysis of observed time series, which is required to train accu-
rate Markov models with short-length time series. In this framework, the underlying stochastic process is assumed to be an ergodic semigroup of endomorphisms (i.e., surjective measure-preserving transformations (MPT)) \cite{36}; and the developed model is generally a time-inhomogeneous Markov chain, which is constructed based on the spectral properties of the underlying physical process as explained below.

The notion of measure-preserving transformation (MPT) has evolved in the discipline of Statistical Mechanics to represent Hamiltonian systems (e.g., Louisville Theorem \cite{37}), where the total energy of the dynamical system is invariant \cite{38}. A key concept in this regard is that, even though a measure-preserving dynamical system is described by a transformation with time-varying eigenvalues, the absolute values of the eigenvectors remain time-invariant under the ergodicity assumption \cite{36}. Based on this rationale, this paper presents a methodology for constructing PFSAs, from short-length time series, which generate non-homogeneous Markov chain models for describing the uncertain dynamics of the physical process. As a result, the time-invariance of the eigenvectors, which reflects measure-invariance of the dynamical system, is used to decide the window length of the time series required to construct PFSAs. Unlike the standard STSA \cite{21, 28} that requires increasing the window length until the resulting PFSA is no longer significantly changing, the proposed MPT-based STSA selects a minimum window length for which the eigenvectors are nearly constant. The rationale is that anomalies are usually associated with a change in the system’s total energy, which naturally makes the system no longer measure-invariant and thus the eigenvectors are no longer time-invariant. Along this line, a metric of variability of eigenvectors is proposed as a measure of anomaly. This paper also develops an MPT-based metric to quantify the divergence of evolving PFSAs from the nominal PFSA; this information is then used for pattern classification and anomaly detection with low-delay tolerance.

The proposed methodology has been validated with experimental data generated from a laboratory apparatus for the aforementioned application, namely, detection of thermoacoustic instabilities (TAI) in combustion systems. The re-
results have been compared with those obtained with a standard STSA (e.g., [28]), which show consistent superiority of the proposed MPT-based STSA.

Contributions: Major contributions of the paper are summarized below:

1. **Spectral analysis for pattern classification and anomaly detection:** The proposed methodology facilitates modeling of uncertain dynamical systems by time-inhomogeneous Markov chains from short-length time series. To this end, STSA-based algorithms have been developed on the spectral property of ergodic semigroups of measure-preserving transformations, where all eigenvectors have time-invariant absolute values [36].

2. **Quantification of evolving anomalies:** Evolving anomalies are quantified as norms of deviations in the eigenvectors of the constructed sequence of stochastic matrices by utilizing the invariant property of the eigenvectors. This norm tends to be small if the dynamical system remains in the nominal state. As the system starts deviating from the nominal state, the absolute values of the eigenvectors no longer remain invariant, and hence the quantified anomaly tends to increase. Diverging patterns are detected as the quantified anomaly exceeds a threshold that is selected by the user during the training phase, based on the receiver operating characteristics.\(^1\)

3. **Validation with experimental data:** The proposed anomaly detection algorithms are validated on a laboratory-scale experimental apparatus for prediction of thermoacoustic instabilities.

Organization: The paper is organized in six sections including the present section. Section 2 briefly provides mathematical preliminaries on measure-preserving transformation and ergodicity, which play a central role in developing the algorithms presented in this paper. Section 3 succinctly presents the background information for STSA. Section 4 presents the technical approach and develops the algorithms. Section 5 presents the results of validation with experimental data.\(^1\)

\(^1\)The Receiver Operating Characteristic (ROC) curves [39] have been used in this paper for assessing the detection performance of MPT-based STSA relative to that of standard STSA.
perimental data. Section 6 summarizes and concludes the paper along with recommendations for future research.

2. Background: Measure-Preserving Transformation and Ergodicity

This section provides a very brief introduction to the notions of measure-preserving transformation and ergodicity, which form the backbone of the statistical pattern classification and anomaly detection methodology presented in this paper; the details are extensively reported by Cornfeld et al. [36]. The following definitions, which are available in standard literature, are presented below for completeness of the paper.

Definition 1. Let $\Omega$ be a nonempty set. A collection $T$ of subsets of $\Omega$ is called a topology and the members of $T$ are called $T$-open (or open) sets provided that the following three conditions are satisfied.

- $\emptyset \in T$ and $\Omega \in T$
- Closure under finite intersection, i.e., $\bigcap_{k=1}^{n} E_k \in T$ if $E_k \in T$ for $k = 1, \cdots, n \ \forall n \in \mathbb{N} \equiv \{1, 2, 3, \cdots\}$.
- Closure under arbitrary (i.e., finite, countable, or uncountable) union, i.e., $\bigcup_{\alpha} E_\alpha \in T$ if $E_\alpha \in T \ \forall \alpha \in I$ that is an arbitrary (i.e., finite, countable, or uncountable) index set.

The pair $(\Omega, T)$ is said to form a topological space, where the complement of an open set is called a closed set, i.e., if $S$ is an open set, then $\Omega \setminus S$ is a closed set.

Definition 2. Let $\Omega$ be a nonempty set. A collection $\mathcal{E}$ of subsets of $\Omega$ is called a $\sigma$-algebra and the members of $\mathcal{E}$ are called $\mathcal{E}$-measurable (or measurable) sets provided that the following three conditions are satisfied.

- $\Omega \in \mathcal{E}$.
- If $E \in \mathcal{E}$, then $\Omega \setminus E \in \mathcal{E}$.
• A countable union of measurable sets is measurable, i.e., if \( \{ E_k \} \) is a countable collection of members of \( \mathcal{E} \), then \( \bigcup_k E_k \in \mathcal{E} \).

The pair \((\Omega, \mathcal{E})\) is said to form a measurable space.

**Definition 3.** Let \((\Omega, \mathcal{E})\) be a measurable space. Then, the value of the set function defined as \( \mu : \mathcal{E} \to [0, 1] \) is called a (probability) measure provided that the following two conditions are satisfied.

\begin{itemize}
  \item \( \mu(\Omega) = 1 \).
  \item If \( \{ E_k \} \) is a countable collection of members of \( \mathcal{E} \), then
    \[ \mu \left( \bigcup_k E_k \right) \leq \sum_k \mu(E_k) \]
    and the equality holds if the members of \( \{ E_k \} \) are pairwise disjoint, i.e., if \( E_i \cap E_j = \emptyset \) \( \forall \ i \neq j \).
\end{itemize}

The triple \((\Omega, \mathcal{E}, \mu)\) is called a measure space; in the present study, the measure space is referred to as probability space because \( \mu(\Omega) = 1 \).

If two measurable sets \( E, F \in \mathcal{E} \) are such that \( \mu(E \Delta F) = 0 \), then it is said that \( E = F \) \( \mu \)-almost everywhere (abbreviated as \( \mu \)-ae). Therefore, all measurable sets that are equal \( \mu \)-ae form an equivalence class; members of this equivalence class are \( \mu \)-almost equal sets. Note: \( \Delta \) is the symmetric difference operation on sets such that \( A \Delta B \equiv (A \setminus B) \cup (B \setminus A) \).

**Definition 4.** Let \((\Omega, \mathcal{E}, \mu)\) be a probability space and let \( T : (\Omega, \mathcal{E}, \mu) \to (\Omega, \mathcal{E}, \mu) \) be a transformation. Then, \( T \) is called measurable if \( T^{-1}E \in \mathcal{E} \) \( \forall E \in \mathcal{E} \).

A measurable set \( E \in \mathcal{E} \) is called \( T \)-invariant if \( \mu(E \Delta T^{-1}E) = 0 \), which implies that \( Tx \in E \) for \( \mu \)-almost all \( x \in E \). Furthermore, a function \( f : \Omega \to [0, \infty) \) is called \( T \)-invariant if \( f(Tx) = f(x) \) for \( \mu \)-almost all \( x \in \Omega \).

A measurable transformation \( T \) is called a measure-preserving transformation (MPT) if \( \mu[T^{-1}E] = \mu(E) \) \( \forall E \in \mathcal{E} \).

A measure-preserving transformation \( T \) is called an endomorphism if \( T \) is surjective (i.e., onto). If \( T \) is bijective (i.e., one-to-one and onto) and preserves measure, then it is called an automorphism.
Remark 1. The concept of MPT has been widely used to investigate the asymptotic properties of random sequences in statistical mechanics [36]. For an endomorphism $T$ on a (finite) measure space $(\Omega, \mathcal{E}, \mu)$, every measurable set $E$ has the recurrence property in the sense that once $E$ is visited, it will be revisited infinitely many times; that is, if $x \in E$, then there are infinitely many values of $n$ such that $T^n x \in E$ [36].

Definition 5. Let $S$ be a (nonempty) set with a binary operation $\circ$. Then, $S$ is called a semigroup if the following two conditions are satisfied:

- **Closure:** $\circ : S \times S \to S$.

- **Associativity:** $(x \circ y) \circ z = x \circ (y \circ z) \ \forall \ x, y, z \in S$.

Definition 6. [40] Let $\{T^n\}$ be a one-parameter semigroup of endomorphisms on a probability space $(\Omega, \mathcal{E}, \mu)$. A function $f \in L^1(\mu)$ is said to be an eigenfunction of $\{T^n\}$ with the eigenvalue $\lambda(n)$ if $f$ is a non-zero function and $f(T^n) = \lambda(n)f$ $\mu$-ae for all $n \in \mathbb{N}$.

The sequence $\{T^n\}$ is said to be ergodic if each $T^n$-invariant set $E \in \mathcal{E}$ is trivial, i.e., either $\mu[E] = 0$ or $\mu[E] = 1 \ \forall n \in \mathbb{N}$.

Remark 2. The concept of ergodicity has been widely used in statistical mechanics and probabilistic modeling of dynamical systems [36]. In an ergodic process, it is sufficient to have a single sufficiently long realization in order to characterize the statistics of the process. Given any discrete-time realization $\{x_n\}$ of an ergodic process $X \in L^1(\mu)$, as $n \to \infty$, the time average $\frac{1}{n} \sum_{j=1}^{n} x_j$ converges $\mu$-ae and in $L^1$ to the ensemble average $\int_{\Omega} X d\mu$ [41].

Another useful formulation of ergodicity is as follows: Given a probability space $(\Omega, \mathcal{E}, \mu)$, the semigroup of endomorphisms $\{T^n\}$ is ergodic if and only if every invariant measurable function is equal to a constant $\mu$-ae on $\Omega$. Based on this formulation, it can be tested whether a semigroup $\{T^n\}$ is ergodic or not by looking at its eigenfunction $f$ corresponding to the eigenvalue $\lambda(n) = 1$, for which $f(x) = f(T^n x)$ for $\mu$-almost all $x \in \Omega$ and for all $n \in \mathbb{N}$. Hence, $f$ is
an invariant function under \( \{T^n\} \) and therefore is a constant \( \mu \)-ae if and only if \( \{T^n\} \) is ergodic. In this context, an ergodic semigroup of endomorphisms has an important property given by the following theorem \cite{36}.

**Theorem 1.** Let \((\Omega, \mathcal{E}, \mu)\) be a probability space, and let \(\{T^\theta\}\) be a semigroup of endomorphisms, where \(\theta \in [0, \infty)\). Then, \(\{T^\theta\}\) is ergodic if and only if the absolute value of every eigenfunction is a constant \(\mu\)-ae. That is, if \(f\) is an eigenfunction of the endomorphism \(T^\theta\), then \(|f(x)|\) is a constant for \(\mu\)-almost all \(x \in \Omega\).

**Proof.** The proof of the theorem is given in \cite{36}.

3. Symbolic Time Series Analysis

Before embarking on a description of the technical approach and the algorithms therein, it is necessary to provide the background for construction of probabilistic finite state automata (PFSA) (see Subsection 3.1 \cite{21}) and \(D\)-Markov machines (see Subsection 3.2 \cite{28}). Subsection 3.1 refers to Section 2 that addresses measure-preserving transformation of ergodic semigroups.

3.1. Probabilistic Finite State Automata

A time series is converted into a symbol string by partitioning the signal space into a finite number of cells, where the number of cells is identically equal to the cardinality \(|\mathcal{A}|\) of the (symbol) alphabet \(\mathcal{A}\), and each cell is assigned one of the symbols in \(\mathcal{A}\). At a given instant of time, a data point is assigned the symbol corresponding to the cell within which the data point is located; details are reported in \cite{28}. The resulting symbol string is used to construct a \(D\)-Markov model, defined in the sequel, that describes the statistics of the underlying stochastic process. The following definitions, which are available in standard literature (e.g., \cite{21, 28}), are recalled here for completeness of the paper.

**Definition 7.** A finite state automaton (FSA) \(G\), having a deterministic algebraic structure, is a triple \((\mathcal{A}, Q, \delta)\) where:
• $\mathcal{A}$ is a (nonempty) finite alphabet, i.e., its cardinality $|\mathcal{A}| \in \mathbb{N}$.

• $\mathcal{Q}$ is a (nonempty) finite set of states, i.e., its cardinality $|\mathcal{Q}| \in \mathbb{N}$.

• $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ is a state transition map.

**Definition 8.** A symbol block, also called a word, is a finite-length string of symbols belonging to the alphabet $\mathcal{A}$, where the length of a word $w \triangleq s_1s_2 \cdots s_\ell$ with $s_i \in \mathcal{A}$ is $|w| = \ell$, and the length of the empty word $\epsilon$ is $|\epsilon| = 0$. The parameters of FSA are extended as:

• The set of all words, constructed from symbols in $\mathcal{A}$ and including the empty word $\epsilon$, is denoted as $\mathcal{A}^*$.

• The set of all words, whose suffix (respectively, prefix) is the word $w$, is denoted as $\mathcal{A}^*w$ (respectively, $w\mathcal{A}^*$).

• The set of all words of (finite) length $\ell$, where $\ell \in \mathbb{N}$, is denoted as $\mathcal{A}^\ell$.

A symbol string (or word) is generated from a (finite-length) time series by symbolization.

**Definition 9.** A probabilistic finite state automaton (PFSA) $\mathcal{K}$ is a pair $(G, \pi)$, where:

• Deterministic FSA $G$ is the underlying algebraic structure of PFSA $\mathcal{K}$.

• The morph function $\pi : \mathcal{Q} \times \mathcal{A} \to [0,1]$ is also known as the symbol generation probability function that satisfies the condition: $\sum_{\sigma \in \mathcal{A}} \pi(q, \sigma) = 1$ for all $q \in \mathcal{Q}$.

Often the state transition probability mass function $\kappa : \mathcal{Q} \times \mathcal{Q} \to [0,1]$ is constructed by combining $\delta$ and $\pi$, which can be structured as a $|\mathcal{Q}| \times |\mathcal{Q}|$ matrix $\mathcal{T}$.

In that case, the PFSA can be described as the triple $\mathcal{K} = (\mathcal{A}, \mathcal{Q}, \mathcal{T})$.

It is noted that the $|\mathcal{Q}| \times |\mathcal{Q}|$ state transition probability matrix $\mathcal{T}$ is stochastic (i.e., each element of $\mathcal{T}$ is non-negative and each row sum is unity). Ergodicity of the underlying process, from which $\mathcal{T}$ is constructed, is equivalent
to irreducibility of $T$, which implies that $T$ has exactly one eigenvalue at unity (i.e., $\lambda = 1$) and that the rest of the eigenvalues are either on or within the unit circle with center at 0 (i.e., $|\lambda| \leq 1$). The (sum-normalized) eigenvector $v^1$ corresponding to the unity eigenvalue (i.e., $\lambda = 1$) represents the stationary state probability vector of the Markov chain.

**Remark 3.** The mathematical concept of ergodic semigroup of endomorphisms and some of the relevant results are presented in Section 3. Following Theorem 3, the absolute value of the eigenfunctions of individual transformations $T^n$ in the one-parameter ergodic semigroup of endomorphisms do not change with $n$ although the respective eigenvalues may vary with $n$. These results are explained below in the context of STSA.

In the probability space $(\Omega, \mathcal{E}, \mu)$ of STSA, $\Omega$ is the (finite) set of states $\Omega$. With symbols $s \in \mathcal{A}$ occurring randomly, the state transition map $\delta : \Omega \times \mathcal{A} \to \Omega$ in Definition 7 becomes a random mapping $T : (\Omega, \mathcal{E}, \mu) \to (\Omega, \mathcal{E}, \mu)$ such that $T(q)$ yields a $\Omega$-valued random variable for each $q \in \Omega$. The state transition probability mass function $\kappa : \Omega \times \Omega \to [0, 1]$ satisfies the following condition:

$$P[T(q) \in E] \triangleq \sum_{\tilde{q} \in E} \kappa(q, \tilde{q}) \quad \forall q \in \Omega \quad \forall E \in \mathcal{E}$$

where $P$ denotes the transition probability with respect to the underlying probability space $(\Omega, \mathcal{E}, P)$, which implies that the random variable $T(q)$ has the probability mass function $\kappa(q, \cdot)$. Then, the $|\Omega| \times |\Omega|$ state transition probability matrix $T$ is a stochastic representation of the random mapping $T$.

**Example 1.** In the probability space $(\Omega, \mathcal{E}, P)$, let $\Omega = \{q_1, q_2\}$ with the $\sigma$-algebra $\mathcal{E} = \{\emptyset, \{q_1\}, \{q_2\}, \Omega\}$, and $P : \mathcal{E} \to [0, 1]$, let $\{T^k\}$ be a sequence of $|\Omega| \times |\Omega|$ stochastic matrices such that $T^k = \begin{bmatrix} p_k & 1 - p_k \\ 1 - p_k & p_k \end{bmatrix}$ where $p_k \in [0, 1)$ can be arbitrary for any given $k$. The eigenvalues of each stochastic matrix $T^k$ are: $\lambda_1 = 1$ and $\lambda_2 = 2p_k - 1$, and the corresponding absolute-sum-normalized left eigenvectors are: $v^1 = [0.5 \ 0.5]$ and $v^2 = [0.5 - 0.5]$ that are $k$-invariant. Since the left eigenvector $v^1 = [0.5 \ 0.5]$, corresponding to the eigenvalue $\lambda_1 = 1$,
is the state probability vector for all $k$ regardless of the value of $p_k$, it follows that, for all $k$, $\mathcal{T}^k$ is measure-preserving because $v^1 \mathcal{T}^k = v^1$, and $\mathcal{T}^k$ is ergodic because the only $\mathcal{T}^k$-invariant measurable sets (i.e., members of $\mathcal{E}$) are $\emptyset$ and $\mathcal{Q}$. It also follows from Definition 4 and Definition 6 that the sequence $\{T^k\}$ of random mappings is measure-preserving and ergodic in this example.

**Definition 10.** A sequence $\{\mathcal{K}^n\} \triangleq \{\mathcal{A}, \mathcal{Q}, \mathcal{T}^n\}$ of PFSA is said to be a semigroup of PFSA if $\mathcal{T}^n \mathcal{T}^m = \mathcal{T}^{n+m}, \forall n, m \in \mathbb{N}$. A semigroup of PFSA $\{\mathcal{K}^n\}$ on a probability space $(\mathcal{Q}, \mathcal{E}, \mu)$ is said to be measure-preserving and ergodic if $\mathcal{T}^n$ is measure-preserving and ergodic $\forall n \in \mathbb{N}$.

A straightforward result of Theorem 1 for the finite-dimensional case is given by the following corollary, which is central to the current paper.

**Corollary 1.** Let $\{\mathcal{K}^n\}$ be a semigroup of measure-preserving PFSA on a probability space $(\mathcal{Q}, \mathcal{E}, \mu)$. Then, $\{\mathcal{K}^n\}$ is ergodic if and only if the absolute value of every eigenvector $v^j(n)$ of $\mathcal{T}^n \forall n \in \mathbb{N}$ and $j = 1, \cdots, M \leq |\mathcal{Q}|$, is uniformly distributed, i.e.,

$$|v^j(n)| = \left[ \frac{1}{|\mathcal{Q}|}, \cdots, \frac{1}{|\mathcal{Q}|} \right]$$

This paper considers a one-parameter semigroup of endomorphisms $\{T^n\}$ of the measure space $(\Omega, \mathcal{E}, \mu)$, i.e., the two properties of closure and associativity are satisfied. That is, $T^{n+m} x = T^n(T^m x)$ for $\mu$-almost all $x \in \Omega$ and all $n, m \in \mathbb{N}$.

### 3.2. D-Markov Machines

When constructing a $D$-Markov machine, it is assumed that the generation of the next symbol depends only on a finite history of at most $D$ consecutive symbols, i.e., a symbol block not exceeding the specified length $D$. In this context, a $D$-Markov machine is defined as follows.

**Definition 11.** A $D$-Markov machine is a PFSA in the sense of Definition 9 and it generates symbols that solely depend on the (most recent) history of
at most $D$ consecutive symbols, where the positive integer $D$ is called the depth of the machine. Equivalently, a $D$-Markov machine is a statistically stationary stochastic process $S = \cdots s_{-1}s_0s_1 \cdots$, where the probability of occurrence of a new symbol depends only on the last consecutive (at most) $D$ symbols, i.e.,

$$P[s_n \mid \cdots s_{n-D} \cdots s_{n-1}] = P[s_n \mid s_{n-D} \cdots s_{n-1}]$$

(1)

Consequently, for $w \in \mathcal{A}^D$ (see Definition 8), the equivalence class $\mathcal{A} \ast w$ of all (finite-length) words, whose suffix is $w$, is qualified to be a $D$-Markov state that is denoted as $w$.

The PFSA of a $D$-Markov machine generates symbol strings including the empty word $\epsilon$. That is, a generated symbol string has the form $\{s_1s_2 \cdots s_\ell\}$, where $s_j \in \mathcal{A}$ and $\ell$ is a positive integer. Both morph function $\pi$ and state probability transition matrix $T$ implicitly support the fact that PFSA satisfy the Markov condition, where generation of a symbol depends only on the current state that is a symbol string of at most $D$ consecutive symbols [28]. For the proposed $D$-Markov-based methodology, there are four primary choices as enumerated below:

1. **Alphabet size ($|\mathcal{A}|$):** Larger is the alphabet size more distinct are the different regimes, but more training data would be needed. There are several algorithms for the choice of optimal alphabet size (e.g., [24]).

2. **Partitioning Method:** While there are many data partitioning techniques, maximum entropy partitioning (MEP) [22, 23, 28] and $K$-means partitioning [43] have been chosen as they are commonly used.

3. **Depth ($D$) in the $D$-Markov machine:** Higher values of the positive integer $D$ may lead to better results at the expense of increased computational time due to larger dimension of the feature space and need for more training. In this paper, $D = 1$ and $D = 2$ have been chosen for data sets.

4. **Choice of Feature:** The feature needs to be one that best captures the nature (e.g., texture) of the signal and is not computationally expensive.
3.3. Anomaly Detection in the STSA Setting

From the perspective of discrete-time measurements, usage of $D$-Markov machines is an efficient and convenient way of modeling a dynamical system. In this setting, the time series $\{x_n\}$, where $x_n \in \mathbb{R}^m$ for some $m \in \mathbb{N}$, is converted into a symbol string $\{s_n\}$, $s_n \in \mathcal{A}$, where $\mathcal{A}$ is a finite-cardinality alphabet. For anomaly detection using the standard STSA methodology, a time series $\{x_n\}$ is first converted into a symbol string. Then, PFSAs are constructed from the symbol strings, which in turn generate low-dimensional feature vectors that are used for detection of anomalous patterns. The procedure is executed in the following steps.

1. Select a block of a time series, called the nominal block, for which the system is in a normal operating condition.
2. Construct a partition for the nominal block and convert it into a symbol string to construct the nominal PFSA model. The emission matrix (and hence the state transition probability matrix) of the PFSA model are constructed by frequency counting. This learned nominal model generates a probability vector $v^0$ that represents the nominal pattern.
3. Select a new block of the time series up to the current time $n$ and convert it into a symbol string using the learned nominal partition. This yields a new PFSA with a new (quasi-)stationary probability vector $v^n$ that represents the feature vector of the system at time $n$.
4. Compute the anomaly at the time epoch $n$ by considering a string of divergences between the nominal feature and current feature vectors and by sliding the block of data $N$ times as:

$$
\nu(n) = \frac{1}{N} \sum_{m=n}^{n+N-1} d_{KL}(v^0, v^m) \tag{2}
$$

where $d_{KL}(\cdot, \cdot)$ is the Kullback-Leibler divergence.

The anomaly metric $\nu(n)$ in Eq. (2) is used in this paper for the standard STSA, which is more robust to outliers and fast fluctuations in the time series than if individual $d_{KL}(v^0, v^m)$ is used.
4. Technical Approach

Let a dynamical system be represented by a semigroup of endomorphisms \( \{T^n\} \) acting on the probability space generated by STSA (see Example 1 in Subsection 3.1). A major issue in the standard STSA-based anomaly detection is that it may require a long time series to construct the stationary state transition probability matrix \( T \), from which the stationary probability vector \( v \) is generated to serve as a feature vector. However, many real-life applications may require low-delay detection of anomalous events, for which short-length time series of measurements must be used. For example, thermoacoustic instabilities (TAI) in combustion systems typically occur on the order of milliseconds, which must be mitigated by fast actuation of control signals. This mandates early detection of instabilities from short-length sensor time series [45].

In view of the above discussion, this paper considers the general case of non-homogeneous Markov-chain models, in which the state transition probability matrix \( T \) is, in general, time-varying. Although the underlying stochastic process could be stationary, short-length time series windows may produce a time-varying \( T \)-matrix; hence, in general, the eigenvalues and eigenvectors of \( T \) are time-varying as well. However, although \( T^n \) and \( T^{n+1} \) might be different, under certain circumstances they would share the same eigenvectors even if the respective eigenvalues are different, as seen in Example 1 (see Subsection 3.1). Moreover, if the matrices \( T^n \) and \( T^{n+1} \) commute and each one of them has \( |Q| \) linearly independent eigenvectors, then they share the same eigenvectors [46]; this issue is further discussed in Section 6 as a topic of future research. It is also seen in Corollary 1 in Subsection 3.1 that if \( \{T^n\} \) is an ergodic semigroup of endomorphisms, then the eigenvectors of \( T^n \) have \( n \)-invariant absolute values. Therefore, by observing \( T^n \) for increasing values of \( n \), the minimum window length \( L \) for STSA (see Section 1 and Algorithm 1) can be determined when this measure-preserving and ergodicity condition is satisfied.

A modification of the standard STSA technique is now proposed based on the variations of eigenvectors. Algorithm 1 presents a procedure for identification
of the window length $L$, where $L$ is kept on increasing until a metric of the variations of the eigenvector, hereafter called ergodicity metric $\rho$, (that should be a constant for measure-preserving and ergodic processes) is less than the (user-specified) threshold parameter $\tau_1$. Algorithm 2 presents the details for detecting anomalous patterns using short-length time series to construct a $D$-Markov machine model for the underlying stochastic process, given the alphabet size $|\mathcal{A}|$, the Markov depth $D$, and window length $L$. In this situation, an anomaly is detected online when the aforementioned ergodicity metric $\rho$ exceeds the (user-specified) threshold parameter $\tau_2$. It is demonstrated in Section 5 that this criterion can be used to achieve class separability in the feature space for pattern classification.

The threshold parameters $\tau_1$ and $\tau_2$ are interdependent in the sense that a smaller $\tau_1$ would result in a larger window length $L$ (i.e., a larger delay tolerance), which would require a smaller $\tau_2$ under a fixed maximum allowable false positive rate (FPR) for similar quality of decisions on anomaly detection; it is noted that both parameters, tolerated delay and maximum allowable FPR, are generally application-dependent. In this regard, Pareto optimization for selection of the threshold parameters $\tau_1$ and $\tau_2$ is recommended as a topic of future research in Section 6.
\begin{algorithm}[H]
\caption{MPT-based selection of window length}
\begin{algorithmic}[1]
\STATE INPUT: Alphabet $\mathcal{A}$, Markov depth $D$, a time series $\{x_n\}$ generated by an ergodic stochastic process, number of sliding windows $N \in \mathbb{N}$, increment of the window length $\Delta L$ in each iteration, and a threshold $\tau_1 > 0$.
\STATE OUTPUT: Window length $L$ of the data blocks $\{x_{n:n+L}\}$ for construction of a $D$-Markov model of the stochastic process.
\STATE 1: Choose an initial window size $L$.
\STATE 2: \textbf{do}
\STATE 3: \hspace{1em} Convert time series blocks $\{x_{n+1:n+L}\}$, $n = 0, 1, \ldots, N - 1$, into symbol strings $\{s_{n+1:n+L}\}$, $s_i \in \mathcal{A}$, using one of the STSA partitioning methods.
\STATE 4: \hspace{1em} Using frequency counting, construct a $D$-Markov machine based on each $s_{n+1:n+L}$ to obtain state transition probability matrices $\{T^n\}$.
\STATE 5: \hspace{1em} Find the left eigenvectors $\{v^i(n)\}$, $i = 1, \ldots, M$, for each one of the state transition probability matrices, where $M$ is the total number of the eigenvectors for each $T^n$. It is noted that $M \leq |Q|$.
\STATE 6: \hspace{1em} $v^i \leftarrow \frac{1}{N} \sum_{n=0}^{N-1} |v^i(n)|$, $i = 1, \ldots, M$, where $|v^i(n)|$ is the vector of the absolute values of the components of $v^i(n)$.
\STATE 7: \hspace{1em} $\theta^i \leftarrow \sum_{n=0}^{N-1} ||v^i(n) - v^n||$, $i = 1, \ldots, M$.
\STATE 8: \hspace{1em} $\rho \leftarrow \max_{i \in \{1, \ldots, M\}} ||\theta^i||_{l_1}$, where $||\theta^i||_{l_1} = \sum_{j=1}^{|Q|} |\theta^i_j|$.\par
\STATE 9: \hspace{1em} $L \leftarrow L + \Delta L$.
\STATE 10: \textbf{while} $\rho > \tau_1$ \textbf{do}
\STATE 11: \hspace{1em} \textbf{end if}
\end{algorithmic}
\end{algorithm}

\begin{algorithm}[H]
\caption{MPT-based anomaly detection}
\begin{algorithmic}[1]
\STATE INPUT: Alphabet $\mathcal{A}$, Markov depth $D$, window length $L$, time series $\{x_n\}$, number of sliding windows $N \in \mathbb{N}$, a threshold $\tau_2 > 0$.
\STATE OUTPUT: The decision on whether the system is nominal or anomalous.
\STATE 1: Convert time series blocks $\{x_{n+1:n+L}\}$, $n = 0, 1, \ldots, N - 1$, into symbol strings $\{s_{n+1:n+L}\}$, $s_i \in \mathcal{A}$, using one of the STSA partitioning methods.
\STATE 2: Using frequency counting, construct a $D$-Markov machine based on each $s_{n+1:n+L}$ to obtain state transition probability matrices $\{T^n\}$.
\STATE 3: Find the left eigenvectors $\{v^i(n)\}$, $i = 1, \ldots, M$, for each one of the state transition probability matrices, where $M$ is the total number of the eigenvectors for each $T^n$. It is noted that $M \leq |Q|$.
\STATE 4: $v^i \leftarrow \frac{1}{N} \sum_{n=0}^{N-1} |v^i(n)|$, $i = 1, \ldots, M$, where $|v^i(n)|$ is the vector of the absolute values of the components of $v^i(n)$.
\STATE 5: $\theta^i \leftarrow \sum_{n=0}^{N-1} ||v^i(n) - v^n||$, $i = 1, \ldots, M$.
\STATE 6: $\rho \leftarrow \max_{i \in \{1, \ldots, M\}} ||\theta^i||_{l_1}$, where $||\theta^i||_{l_1} = \sum_{j=1}^{|Q|} |\theta^i_j|$.
\STATE 7: \textbf{if} $\rho > \tau_2$ \textbf{then}
\STATE 8: \hspace{1em} declare the system as anomalous\par
\STATE 9: \textbf{else}
\STATE 10: \hspace{1em} declare the system as nominal\par
\STATE 11: \textbf{end if}
\end{algorithmic}
\end{algorithm}
Remark 4. Although Algorithms 1 and 2 theoretically require $\theta^i$’s for all $i = 1, \cdots, M$, it may often turn out that it is necessary to check only one of these parameters. It is demonstrated in the specific example considered in Section 6 that non-ergodicity of a stochastic process may cause all eigenvectors to have time-varying absolute values. In this case, it may be sufficient to look at one of these eigenvectors to check for measure preservation and ergodicity. However, depending on the particular application, some of these eigenvectors could be more information-bearing than the others for discrimination of anomalous behavior from the nominal, in which case the best eigenvector is chosen based on a held-out training data set. Instead of using all components of the eigenvector $v^i(n)$ for computing $\theta^i$ in Algorithms 1 and 2, one may use only some of these components. Alternatively, one may use $||v^i(n)||_\infty$ instead of $v^i(n)$ for computing $\theta^i$; then, $\theta^i$ becomes a scalar and computations are further reduced.

5. Experimental Validation: Results & Discussion

This section validates the MPT-based STSA methodology (see Section 4) on experimental data, generated from a laboratory apparatus for predicting thermoacoustic instabilities (TAI) in combustion systems. The objective here is to evaluate the performance of the proposed MPT-based STSA, presented in Algorithms 1 and 2 for anomaly detection using short-length time series and low-dimensional feature vectors, and its robustness to changes in the parameters of data partitioning and the algorithms. In this regard, performance of MPT-based STSA is compared with that of standard STSA (e.g., [28]) (see Subsection 3.3) on the same experimental data. Table 1 lists the major differences, demonstrated in the sequel, between standard STSA and MPT-based STSA.

5.1. Description of the Laboratory Apparatus and Experimental Data

Figure 1 depicts the experimental apparatus for emulation of thermoacoustic instabilities (TAI) on an electrically heated Rijke tube apparatus [47], which has been commonly used by researchers for studying TAI, because electrical heating is easier to operate in the laboratory environment than fuel burning, and yet
Table 1: Comparison of standard STSA and MPT-based STSA

<table>
<thead>
<tr>
<th></th>
<th>Standard STSA</th>
<th>MPT-based STSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Time-invariance of nominal-phase PFSA</td>
<td>Time-invariance of absolute values of eigenvectors of nominal-phase PFSA that are time-varying, in general.</td>
</tr>
<tr>
<td>2</td>
<td>Generation of homogeneous Markov-chain models</td>
<td>Generation of non-homogeneous Markov-chain models</td>
</tr>
<tr>
<td>3</td>
<td>Anomaly quantification by divergence of the current PFSA from the nominal PFSA</td>
<td>Anomaly quantification by variability of eigenvectors of evolving PFSA</td>
</tr>
<tr>
<td>4</td>
<td>Requirement of relatively long time series for modeling of the underlying process dynamics</td>
<td>Requirement of relatively short time series for modeling of the underlying process dynamics</td>
</tr>
<tr>
<td>5</td>
<td>Less robust to parametric changes in data partitioning and detection system (e.g., alphabet size $</td>
<td>\mathcal{A}</td>
</tr>
</tbody>
</table>

Figure 1: The electrically heated Rijke tube apparatus.

It can emulate the salient properties of TAI in real-life combustors [48]. The Rijke tube in Figure 1 consists of a 1.5 m long horizontal tube with an external cross-section of $10\text{cm} \times 10\text{cm}$ with a wall thickness of $6.5\text{mm}$. It is equipped with an air-flow controller that regulates the flow of air at atmospheric pressure through the tube. It has a heating element placed at quarter length of the tube from the air-input end. A programmable DC power supply controls the power input to the heater. The experiments were conducted in the following manner.

1. For every run, the air flow-rate was set at a constant value. Different runs were performed with flow-rates ranging from 130 liters per minute (LPM)
to 250 LPM at intervals of 20 LPM.
2. First the Rijke tube system was heated to a steady state with a primary heater power input of $\sim 200 \text{ W}$.
3. Then the power input was abruptly increased to a higher value that shows limit cycle behavior as depicted in the stability chart in Mondal et al. [?].

Fifteen (15) experiments were conducted on the Rijke tube apparatus, where the process started with the nominal or stable behavior, and gradually became anomalous and eventually unstable. A time series of pressure oscillations was collected over 30 sec for each experiment, sampled at 8192 Hz, and filtered to attenuate the effects of low-frequency environmental acoustics; typical profiles of the pressure time series are presented in Figure 2. In the detection phase, each pressure signal is truncated at the end of the transient phase so that the pressure signal has only two parts; stable and unstable.

![Unsteady pressure signals showing the transition from stable (nominal) to unstable limit cycle (anomalous) behavior: (a) input power abruptly increased to 1800 W with air flow rate 210 LPM; (b) input power abruptly increased to 2000 with air flow rate 250 LPM.](image)

When the signal space is partitioned in the stable part, the resulting symbol sequences can be described by a (generally) time-varying, measure-preserving and ergodic PFSA, where there is no significant energy change in the system, and each state is expected to be revisited infinitely many times, assuming that the system stays ergodic (i.e., stable) forever. Once the transient phase starts, more energy is added to the system, and some of the PFSA state may no longer be revisited many times. Therefore, in the transient phase, both measure-
preserving and ergodicity properties are lost.

Figure 3: For the MPT-based STSA, convergence of left eigenvector $v^1 = [v_1^1 1 - v_1^1]$, corresponding to the eigenvalue ($\lambda_1 = 1$), to a constant in the stable combustion phase, and to an oscillating function in the transient to unstable phase ($|A| = 2$, $D = 1$).

5.2. Results of Data Analysis

An ensemble of pressure time series have been generated for the nominal (stable) operation and for the anomalous (transient to unstable) operation. To reduce computations, only one of the eigenvectors of the state transition probability matrix $T$ is used (see Remark 4). In this work, the left eigenvector, $v^1$, of $T$, corresponding to the eigenvalue $\lambda_1 = 1$, has been used as the feature vector. The effects of different values of window length $L$ are investigated in three plates of Figure 3 while the other two parameters are set at $|A| = 2$ and $D = 1$. Therefore, the number of $D$-Markov states is $|Q| = 2$ (see Section 3); consequently, the (sum-normalized) eigenvector is 2-dimensional in the form of $v^1(n) = [v_1^1(n) 1 - v_1^1(n)]$ and only the (scalar) values of $v_1^1(n)$ are plotted.

Figure 3 shows that how the absolute value of the left eigenvector (corresponding to $\lambda_1 = 1$) shrinks to a nearly constant vector of value $[0.5 \ 0.5]$ for the stable phase as the window size $L$ is increased, while retaining the time-varying characteristics of the transient zone. In view of Corollary 4 in Subsection 3.1, this observation supports the postulation that the stable phase is considered to be ergodic while the transient phase as non-ergodic. The unstable zones are

---

2In this paper, eigenvectors are always absolute-sum-normalized; therefore, given a $|Q|$-dimensional eigenvector, only $|Q| - 1$ elements of the eigenvector need to be specified.
ignored in Figure 3, because the main objective here is to detect the onset of transience from the stable to the unstable phase.

Figure 4: For MPT-based STSA, convergence of left eigenvector \( v^1 = [v^1_1 \ v^1_2 \ 1 - v^1_1 - v^1_2] \), corresponding to the eigenvalue \( \lambda = 1 \), to the uniform distribution \([1/3 \ 1/3 \ 1/3]\) in the stable combustion phase (\(|A| = 3, D = 1\)).

Figure 4 shows the shrinking behavior of the absolute value of the eigenvectors (corresponding to \( \lambda_1 = 1 \)) with \(|A| = 3, D = 1\), and different window length \( L \)’s. It follows from Figure 4 that the absolute value of the eigenvectors shrink to a nearly constant vector of value \([1/3 \ 1/3 \ 1/3]\) as the window length \( L \) is increased, which is supported by Corollary 1 in Subsection 3.1.

Each of the twelve plates in Figure 5 shows the behavior of the anomaly metric \( \rho(n) \), proposed in Algorithms 1 and 2, over different values of window length \( L \). In this case too, only the eigenvectors corresponding to the eigenvalue \( \lambda_1 = 1 \), are considered. Each plate shows how this metric can efficiently achieve excellent class separability for a reasonable value of the window length \( L \). In this case, Class 1 represents the stable phase and Class 2 represents the transient phase. As shown in the plates of Figure 5 for small values of \( L \), the metric \( \rho(n) \) changes with time for both stable and transient phases. When \( L \) is increased, the metric \( \rho(n) \) starts shrinking in the stable (ergodic) phase, while it keeps on changing in the transient (non-ergodic) phase. Eventually \( \rho(n) \) provides complete class separability at \( L = 240 \). This observation demonstrates the power of the metric \( \rho \) to achieve accurate pattern classification and hence detection of TAI onset in combustion systems.

Algorithm 1 is validated in Figure 6, where the ergodicity metric \( \rho \) in Algorithm 1 is evaluated for different values of \( L \) for a pressure time series from
Figure 5: Class separation in MPT-based STSA with eigenvectors corresponding to the eigenvalue $\lambda_1 = 1$ and using the ergodicity metric $\rho$ in Algorithm 2 ($|\mathcal{A}| = 2$, $D = 1$, and different window length $L$). [Abscissa: Time step $n$; Ordinate: Ergodicity metric $\rho(n)$]

Note: Class 1 $\equiv$ Stable Phase; Class 2 $\equiv$ Transient Phase.

a typical experiment, and the area under the curve (AUC) of the ROC curve is obtained by using Algorithm 2 for the same experiment and the same values of $L$, with parameters $|\mathcal{A}| = 2$, $D = 1$, and K-means for partitioning. As seen in plates (a) and (b) of Figure 6, the profile of ROC performance (i.e., AUC) is strongly correlated to that of the ergodicity metric $\rho$ in Algorithm 4.
Figure 6: Comparison of convergence of MPT-based STSA (using eigenvectors corresponding to the eigenvalue \( \lambda_1 = 1 \)) to that of standard STSA (K-means is used for partitioning in both methods with \( |\mathcal{X}| = 2 \) and \( D = 1 \)): (a) convergence of ergodicity metric \( \rho \) as \( L \) increases (see Algorithm 1) for a typical time series, (b) increase of area under the ROC curve, using Algorithm 2 as \( L \) increases for the same time series, and (c) convergence of the difference norm of consecutive state transition probability matrices for the same time series.

It can be concluded that smaller the value of \( \rho \) more accurate is the \( D \)-Markov model of the stochastic process and hence better ROC is achieved by the \( D \)-Markov model. This conclusion demonstrates the efficacy of Algorithm 1 for selection of the window length \( L \) based on the value of the ergodicity metric \( \rho \). In practice, one may use a number of time series from several experiments to construct a family of plots. Then, the threshold parameter \( \tau_1 \) in Algorithm 1 can be identified by specifying the desired performance in terms of the areas under the ROC curves. That is, the value of \( \tau_1 \) can be chosen such that the corresponding area under the ROC curve for each experiment (or the average area of the ROC curves) is equal to or greater than the desired AUC. The two plates (a) and (c) of Figure 6 show a comparison between the convergence of a sequence of state transition probability matrices \( \{ \mathcal{T}^n \} \) and the convergence of the ergodicity metric \( \rho \), proposed in Algorithm 1 as the window length \( L \) is increased. Again, only the eigenvector corresponding to \( \lambda_1 = 1 \) is considered in computing \( \rho \). As seen from plates (a) and (c) of Figure 6, the convergence of \( \rho \) is much faster and also more smooth than that of the differences in consecutive state transition probability matrices \( \mathcal{T} \). In summary, Figure 6 suggests that the window length \( L \) required for constructing an accurate time-inhomogeneous Markov model based on MPT-based STSA is much less than that required for...
construction of an accurate time-homogeneous Markov model using standard STSA in which the state transition probability matrix is stationary [21].

![AUC: Standard STSA Method = 0.66249, Ergodicity-based Method = 0.98907](image1)

(a) \( D = 1, \ L = 50 \)

![AUC: Standard STSA Method = 0.69389, Ergodicity-based Method = 0.99475](image2)

(b) \( D = 1, \ L = 200 \)

![AUC: Standard STSA Method = 0.56354, Ergodicity-based Method = 0.98999](image3)

(c) \( D = 2, \ L = 200 \)

Figure 7: Comparison of ROC performance: MPT-based STSA and standard STSA (both having K-means partitioning (|\( \mathcal{A} \)| = 2) and eigenvectors with the eigenvalue \( \lambda_1 = 1 \)).

![AUC: Standard STSA Method = 0.82119, Ergodicity-based Method = 0.96109](image4)

(a) \( D = 1, \ L = 50 \)

![AUC: Standard STSA Method = 0.94109, Ergodicity-based Method = 0.99991](image5)

(b) \( D = 1, \ L = 200 \)

![AUC: Standard STSA Method = 0.92932, Ergodicity-based Method = 0.99960](image6)

(c) \( D = 2, \ L = 200 \)

Figure 8: Comparison of ROC performance: MPT-based STSA and standard STSA (both having MEP partitioning (|\( \mathcal{A} \)| = 2) and eigenvectors with the eigenvalue \( \lambda_1 = 1 \)).

Figure 7 shows the ROC curve generated by Algorithm 2 compared to the ROC curve generated by the standard STSA, with K-means used for partitioning, alphabet size \( |\mathcal{A}| = 2 \), and three different pairs of the Markov depth \( D \) and window size \( L \). As seen in Figure 7, MPT-based STSA achieves excellent performance for all three cases, while standard STSA yields significantly inferior performance possibly due to small window lengths and small alphabet size.

In order to improve the performance of the standard STSA, maximum entropy partitioning (MEP) [22] is now used instead of K-means [43]. The comparison results are shown in Figure 8 for the same values of alphabet size \( |\mathcal{A}| \), Markov depth \( D \), and window lengths \( L \). The figure shows that MEP yields significant improvement for standard STSA. However, MPT-based STSA still achieves a significantly better performance for the same parameters and the
same partitioning method. This also demonstrates that MPT-based STSA is more robust to different choices of the partitioning method than standard STSA.

As explained in Remark 4 at the end of Section 4, one may reduce computations by considering only some of the eigenvector’s components rather than all the components. Therefore, only the first component of each eigenvector is considered and the rest of the $(|\Omega| - 1)$ components (only if $|\Omega| > 2$) are ignored in the detection process using MPT-based STSA.

5.3. Selection of the threshold parameters $\tau_1$ and $\tau_2$ in Algorithms 1 and 2

This subsection suggests a procedure for selecting the threshold parameters $\tau_1$ and $\tau_2$ in Algorithms 1 and 2 from the receiver operator characteristic (ROC) curves. Rigorous statistical analysis in this direction is suggested as a topic of future research in Section 6.

The ROC curves generated by Algorithm 2 in Figure 7 and Figure 8 have been generated by varying the threshold parameter $\tau_2$, where each point in the ROC curve corresponds to a specific value of $\tau_2$. Therefore, $\tau_2$ can be selected from a ROC curve by specifying a maximum allowable false positive rate (FPR), which may depend on the application and the user. If the cost of a positive false alarm is low, the maximum FPR could be increased. On the other hand, for applications where the cost for a positive false alarm is high, a small value for the maximum FPR should be selected. The value of the threshold $\tau_1$ in Algorithm 1 can be decided by specifying a maximum delay determined by the window size $L$.

<table>
<thead>
<tr>
<th>Parameters &amp; partitioning method</th>
<th>Standard STSA</th>
<th>MPT-based STSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>A</td>
<td>= 2, D = 1, L = 50, K-means$</td>
</tr>
<tr>
<td>$</td>
<td>A</td>
<td>= 2, D = 1, L = 200, K-means$</td>
</tr>
<tr>
<td>$</td>
<td>A</td>
<td>= 2, D = 2, L = 200, K-means$</td>
</tr>
<tr>
<td>$</td>
<td>A</td>
<td>= 2, D = 1, L = 50, MEP$</td>
</tr>
<tr>
<td>$</td>
<td>A</td>
<td>= 2, D = 1, L = 200, MEP$</td>
</tr>
<tr>
<td>$</td>
<td>A</td>
<td>= 2, D = 2, L = 200, MEP$</td>
</tr>
</tbody>
</table>
5.4. Computational Time for Execution of STSA

Table 2 lists the statistics of CPU execution time of both standard STSA and MPT-based STSA for TAI detection using an ensemble of time series data. It is observed in Table 2 that the execution time of standard STSA is modestly less than that of MPT-based STSA for the same data and the same values of alphabet size $|\mathcal{A}|$, Markov depth $D$ and window length $L$. However, as seen from the detection performance in the ROC curves of Figures 7 and 8, MPT-based STSA requires much shorter lengths ($L$) of time series to achieve similar performance in comparison to standard STSA. This implies that MPT-based STSA would be subjected to much less time delay, compared to standard STSA, to acquire the requisite data for online detection. Table 2 also shows that as the window length $L$ is increased, the standard deviation of the CPU execution time tends to increase as well. The rationale is that, as $L$ is increased, the number of sliding windows for the same time series would decrease resulting in a higher standard deviation.

6. Summary, Conclusion and Future Work

This paper has presented and validated a symbolic time series analysis (STSA) methodology for pattern classification and online anomaly detection in uncertain dynamical systems, where the underlying concept is built upon spectral analysis of ergodic semigroup of measure-preserving transformations. Rather than constructing a homogeneous (i.e., time-invariant) Markov-chain model, which may require a longer window of time series, a non-homogeneous Markov-chain model is constructed, for which consecutive stochastic matrices might be different but they share nearly the same absolute values of the eigenvectors’ components. Based on this information, Markov-chain models are constructed using short windows of time series to facilitate online anomaly detec-

---

3The experimental results were generated on a DELL PRECISION T3400, with an Intel(R) Core(TM)2 Quad CPU Q9550 at 2.83 GHz, with 8 GB RAM, running under Windows 10.
tion. The proposed methodology of pattern classification and anomaly detection, called MPT-based STSA, has been validated by comparison with standard STSA (e.g., [28]) (see Subsection 3.3) on experimental data generated from a laboratory apparatus for detection of thermoacoustic instabilities in combustion systems. In this application, MPT-based STSA shows consistently superior performance compared to the standard STSA.

While there are many areas of theoretical and experimental research to enhance the work reported in this paper, the authors suggest the following topics for future research:

1. Enhancement of computational efficiency by using a commutator operator on ergodic matrices instead of the norm of eigenvector difference. (See the second paragraph of Section 4.)

2. Rigorous statistical analysis (e.g., [3]) and Pareto optimization for selection of the threshold parameters $\tau_1$ and $\tau_2$ in Algorithms 1 and 2, respectively. (See the last paragraph of Section 4.)

3. Experimental validation of MPT-based STSA in diverse applications for pattern classification and anomaly detection.

Acknowledgements

The authors are grateful to Mr. Sudeepta Mondal who has kindly provided the experimental data on the Rijke tube apparatus. The work reported in this paper has been supported in part by U.S. Air Force Office of Scientific Research (AFOSR) under Grant Nos. FA9550-15-1-0400 and FA9550-18-1-0135 in the area of dynamic data-driven application systems (DDDAS). Any opinions, findings, and conclusions in this paper are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

References


[34] V. A. Rohlin, Generators in ergodic theory, Vestnik Leningradskogo Universiteta 18 (1) (1963) 26–32.


