Symbolic Time Series Analysis for Anomaly Detection in Measure-Invariant Ergodic Systems

Najah F. Ghalyan
Research Scholar in Mechanical Engineering
The Pennsylvania State University
University Park, PA 16802
Email: nlfjasim76@gmail.com

Asok Ray, Fellow ASME
Distinguished Professor of Mechanical Engineering and Mathematics
The Pennsylvania State University
University Park, PA 16802
Email: axr2@psu.edu

This paper presents a novel framework of symbolic time series analysis (STSA) for anomaly detection in dynamical systems, where the core concept is built upon a property of ergodic semigroups of measure-preserving transformations (MPTs). Under the ergodicity assumption, the eigenfunctions of these measure-preserving transformations would have time-invariant absolute values. As a result, unlike a standard STSA that is required to generate time-homogeneous Markov chains, the proposed MPT-based STSA is allowed to have time-inhomogeneous Markov chains, where the (possibly time-varying) state transition probability matrix has time-invariant absolute values of eigenvectors. Such a time-invariance facilitates analysis of the dynamical system by using short-length time series of measurements. This is particularly important in applications, where the underlying dynamics and process anomalies need fast monitoring & control actions in order to mitigate any potential structural damage and/or to avoid a catastrophic failure. The proposed MPT-based STSA has been applied for low-delay detection of fatigue damage, which is a common source of failure in mechanical structures and which is known to have uncertain dynamical characteristics. For such applications, a novel anomaly-detection method has been developed, whose performance is validated with experimental data generated from a laboratory apparatus that uses ultrasonic sensor measurements to detect fatigue damage in polycrystalline metallic alloys. Performance of the proposed MPT-based STSA is evaluated by comparison with those of a standard STSA and a hidden Markov model (HMM) on the same experimental data. The results consistently show superior performance of the MPT-based STSA.

Keywords: Anomaly Detection; Ergodicity; Fatigue Damage; Measure Invariance; Spectral Analysis; Symbolic Dynamics.

1 Introduction

Data-driven detection of anomalous behavior [1] is important in diverse engineering applications (e.g., prediction of fatigue failures, condition-based maintenance, and radar systems) as well as for DNA analysis and malware detection (e.g., [2] and references therein). One of the commonly used methods for anomaly detection is the cumulative sum (CUSUM) technique, developed by Page [3], which has been widely used for detection of change points [4]. This tool is very efficient for detecting changes in the time series, which may occur as a result of abrupt variations in the underlying model structure. However, these changes in the time series may happen due to gradual degradation in the underlying dynamical system. Detection of such changes is more challenging than those due to abrupt changes.

Hidden Markov models (HMMs) have been widely used for both change point and anomaly detection, such as speech recognition [5] and early detection of thermo-acoustic instability in combustion systems [6, 7]. In this setting, the (nominal) HMM is trained by using observed time series that is known to represent the nominal behavior. Then, if a change occurs in the time series, the likelihood of the new observed subsequence under the nominal HMM is expected to deviate significantly from the nominal likelihood [8].

In a similar context, the concept of symbolic time series analysis (STSA) has been used by many researchers (e.g., [9, 10, 11]) for constructing Markov-chain models from the respective observed time series. In the STSA framework, a (finite-length) time series is partitioned for conversion into a string of symbols from a (finite-cardinality) alphabet of symbols (e.g., [12, 13, 14, 15]). Subsequently, a probabilistic finite state automaton (PFSA) is constructed from the symbol string (e.g., [16, 17, 18]), in which the probability distribution of the emitted symbols depends upon the immediately
preceeding $D$ symbols, where the Markov depth $D$ is a positive integer. Such a PFSA is called a D-Markov machine, which has found diverse applications in anomaly detection and pattern classification (e.g., [11][12][18][19]). The main distinction between HMMs and D-Markov machines is that the state transition in HMMs could have a non-deterministic algebraic structure [17], which requires an iterative method (e.g., the Baum-Welch algorithm [5]) for training the HMM parameters; in this scenario, the algorithm might lead to a poor local optimum. In contrast, D-Markov machines have a deterministic algebraic structure [18], which makes computation much simpler and less prone to local optimality issues.

In the STSA setting, the selection of the window length of time series to construct the PFSA largely depends on several parameters (e.g., Markov depth and alphabet size) and the nature of the particular underlying process that generates the time series [18]. 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 lengths. Under the assumption of statistical stationarity [11], the (possibly time-varying) state transition probability matrix converges to a constant matrix when the window length may become arbitrarily large [11]. Thus, one may choose a minimum window length at which this matrix tends to be approximately time-invariant. The resulting model in this case would be a time-homogeneous Markov chain [20]. However, this scenario would typically require a large window length, which could be infeasible in many applications where a decision needs to be made with low delay-tolerance.

The notion of measure-preserving transformation (MPT) has been widely used to represent Hamiltonian (i.e., conservative) dynamical systems evolving on a probability space such that the total energy of the dynamical system is constant [21][22]. A key concept in this regard is that even though a measure-preserving dynamical system could be described by a semigroup of transformations with time-varying eigenvalues, the absolute values of the eigenfunctions remain unchanged with time under the ergodicity assumption. Based on this rationale, this paper presents a methodology for constructing PFSA, from short-length time series, which may generate a non-homogeneous Markov chain model for adequately describing the underlying stochastic process. As a result, the time-invariance of the absolute values of eigenvectors, which reflects measure-invariance of the underlying dynamical system, can be used to decide the window length of the time series required to construct the PFSA. Unlike the standard STSA [11][18] that requires increasing the window length until the resulting PFSA is no longer significantly changing, the proposed MPT-based STSA increases the window length until 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 absolute values of eigenvectors are no longer time-invariant. Along this line, a metric of variability in the absolute values of eigenvectors is proposed as a measure of anomaly. The application example in this paper focuses on fatigue damage in polycrystalline metallic alloys (e.g., [23]), which is a common source of failure in mechanical structures. The performance of the proposed MPT-based STSA is evaluated by comparison with those of a standard STSA and an HMM on the same experimental data sets. These data sets have been generated from a laboratory apparatus that uses ultrasonic sensor measurements to detect fatigue damage in polycrystalline metallic alloys.

Major contributions of the paper:
1. Construction of an MPT-based framework of STSA for non-homogeneous Markov chain modeling: The underlying algorithms are built upon the concept of measure-invariance in dynamical systems [22][24][25], which facilitates ergodic measure-preserving modeling of dynamical systems from short-length time series.

2. Identification of a metric for machine/process anomaly: Evolving anomalies are quantified as a norm of deviations in the eigenvectors of the constructed sequence of stochastic matrices by utilizing the invariant property of absolute values of 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 constant, and hence the quantified anomaly tends to increase. Anomalous patterns are detected as the metric exceeds a user-selected threshold.

3. Validation with experimental data: The proposed anomaly detection methodology is validated on a laboratory-scale experimental apparatus for detection of fatigue damage in polycrystalline metallic alloys.

Organization of the paper: The paper is organized in six sections, including the present one. Section 2 provides background information on measure-preservation and ergodicity in dynamical systems. Section 3 briefly describes the principle of STSA for anomaly detection. Section 4 presents the technical approach for developing STSA-based anomaly detection algorithms that rely on spectral properties of semigroups of MPTs. Section 5 presents the results of validation with experimental data. Section 6 summarizes and concludes the paper along with recommendations for future research.

2 Measure-preserving Transformation and Ergodicity

This section provides a brief introduction to the notion of measure-preserving transformation (MPT) that forms the backbone of the methodology presented in this paper; the details are extensively reported in literature (e.g., [26]). The following definitions are presented below for completeness of the current paper and ease of readability.

Definition 1. 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 2.** Let \((\Omega, \mathcal{E})\) be a measurable space. Then, the value of the set function, defined as \(P : \mathcal{E} \to [0, 1]\), is called a probability measure provided that the following two conditions are satisfied:

- \(P[\Omega] = 1\).
- If \(\{E_k\}\) is a disjoint countable collection of members of \(\mathcal{E}\), then \(P\left[\bigcup_k E_k\right] = \sum_k P[E_k]\).

The triple \((\Omega, \mathcal{E}, P)\) is called a probability space.

If two measurable sets \(E, F \in \mathcal{E}\) are such that \(P[E \Delta F] = 0\), then it is said that \(E = F\) \(P\)-almost everywhere (abbreviated as \(P\)-ae) or for \(P\)-almost all \(x \in \Omega\). Therefore, all measurable sets that are equal \(P\)-ae form an equivalence class; members of this equivalence class are \(P\)-almost equal sets. Note: The symmetric difference \((E \Delta F) = (E \setminus F) \cup (F \setminus E)\).

**Definition 3.** Let \((\Omega, \mathcal{E}, P)\) be a probability space and let \(T : (\Omega, \mathcal{E}, P) \to (\Omega, \mathcal{E}, P)\) be a transformation. Then, \(T\) is called measurable if \(T^{-1}E \in \mathcal{E} \quad \forall E \in \mathcal{E}\).

A measurable set \(E \in \mathcal{E}\) is called \(T\)-invariant if \(P[T^{-1}E] = P[E] \quad \forall E \in \mathcal{E}\), which implies that \(T x \in E\) for \(P\)-almost all \(x \in E\). Furthermore, a function \(f : \Omega \to \mathbb{C}\) is called \(T\)-invariant if \(f(T x) = f(x)\) for \(P\)-almost all \(x \in \Omega\), where \(\mathbb{C}\) is the field of complex numbers.

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

**Remark 1.** The concept of MPT has been widely used to investigate the asymptotic properties of random sequences in statistical mechanics \([26]\). For an MPT \(T\) on a (finite) measure space \((\Omega, \mathcal{E}, P)\), every measurable set \(E\) has a 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 \(T^n x \in E [26]\).

**Definition 4.** \([27]\) Let \(\{T^n\}\) be a one-parameter semigroup of MPTs on a probability space \((\Omega, \mathcal{E}, P)\). Note: An algebraic system \((S, \circ)\) is called a semigroup if the following two conditions hold: (i) closure, i.e., \(\circ : S \times S \to S\) and (ii) associativity, i.e., \(\circ(\circ(x, y), z) = \circ(x, \circ(y, z)) \quad \forall x, y, z \in S\).

A function \(f \in L_1(\mathbb{P})\) is said to be an eigenfunction of a sequence \(\{T^n\}\) of transformations with the corresponding sequence \(\{\lambda_n\}\) of (scalar) eigenvalues if \(f\) is a non-zero function such that \(f(T^n) = \lambda_n f\) \(P\)-ae and \(\forall n \in \mathbb{N}\).

The sequence \(\{T^n\}\) is said to be ergodic and \(P\) is called an ergodic measure if each \(T^n\)-invariant set \(E \in \mathcal{E}\) is trivial, i.e., either \(P[E] = 0\) or \(P[E] = 1\).

The following theorem provides a property of MPTs on a probability space \((\Omega, \mathcal{E}, P)\) with an ergodic measure \(P\).

**Theorem 2.** \([26]\) Let \((\Omega, \mathcal{E}, P)\) be a probability space, and let \(\{T^n\}\) be a semigroup of MPTs, where \(t \in [0, \infty)\). Then, \(\{T^n\}\) is ergodic if and only if the absolute value of every eigenfunction is a constant \(P\)-ae. That is, if \(f\) is an eigenfunction of the MPT sequence \(\{T^n\}\), then \(|f(x)|\) is a constant for \(P\)-almost all \(x \in \Omega\).

**Proof.** The proof of the theorem is given in \([25]\).
Remark 2. It is noted that the $|Ω| \times |Ω| \times |Ω|$ state transition probability matrix $T$ is stochastic (28) (i.e., each element of $T$ is non-negative and each row sum is unity). Ergodicity of the underlying process, from which $T$ is reconstructed, is equivalent to irreducibility of $T$ (29), which implies that $T$ has exactly one eigenvalue at unity (i.e., $λ = 1$) and that the rest of the eigenvalues are either on or within the unit circle with center at 0 (i.e., $|λ| ≤ 1$). The (sum-normalized) left eigenvector $v$ corresponding to the unity eigenvalue (i.e., $λ = 1$) represents the stationary state probability vector of the Markov chain (28).

The mathematical concept of ergodic semigroup of endomorphisms and some of the relevant results have been presented in Section 2. Following Theorem 2.1, the absolute-sum-normalized left eigenvalues may vary with $n$ although the respective eigenvalues may vary with $n$. These results are explained below in the context of STSA.

In the probability space $(Ω, ε, P)$ for STSA, the sample space $Ω$ is the (finite) state space $Ω$ of the PFSA under consideration, the associated $σ$-algebra $ε$ is the power set $2^Ω$, and $P$ is the probability measure (see Section 2). The objective here is to model the system dynamics from a time series, $x_n$, of measurements for anomaly detection in the STSA setting. With symbols $s ∈ Α$ occurring randomly, the state transition map $δ : Α \times Α → Α$ in Definition 5 becomes a random mapping $T : (Ω, ε, P) → (Ω, ε, P)$ such that $T(q)$ yields a $Ω$-valued random variable for each $q ∈ Ω$. The state transition probability mass function $κ : Ω × Ω → [0, 1]$ satisfies the following condition:

$$P[T(q)] ∈ 2^Ω \triangleq \sum_{q ∈ Ω} κ(q, ̅q) \quad ∀q ∈ Ω \quad (1)$$

where the state transition probability $κ(q, ̅q)$ is computed with respect to the underlying probability space $(Ω, ε, P)$, which implies that the random variable $T(q)$ has the probability mass function $κ(q, ̅q)$. Note: The $|Ω| × |Ω|$ state transition probability matrix $T$ in Definition 5 is a stochastic representation of the random mapping $T$ (see Remark 2).

Example 1. In the probability space $(Ω, ε, P)$, let $Ω = \{q_1, q_2\}$ with the $σ$-algebra $ε = 2^Ω \triangleq \{\emptyset, \{q_1\}, \{q_2\}, Ω\}$, and $P : ε → [0, 1]$, let $T^k$ be a sequence of $|Ω| \times |Ω|$ 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: $λ_1 = 1$ and $λ_2 = 2p_k − 1$, and the corresponding absolute-sum-normalized left eigenvectors are: $v_1 = [0.5, 0.5]$ and $v_2 = [0.5, −0.5]$ that are $κ$-invariant. Since the left eigenvector $v_1 = [0.5, 0.5]$, corresponding to the eigenvalue $λ_1 = 1$, is the stationary state probability vector for all $k$ regardless of the value of $p_k$, it follows that, for all $k$, $T^k$ is measure-preserving because $v_1 T^k = v_1$. Moreover, $T^k$ is ergodic because $T^k$ is a stochastic representation of the random mapping $T^k$, and the only $T^k$-invariant measurable sets (i.e., members of $ε$) are $\emptyset$ and $Ω$. It follows from Definitions 3 and 4 in Section 2 that the sequence $\{T^k\}$ is measure-preserving and ergodic in this example.

Following Definition 4, a sequence $\{R^n\} \triangleq \{(Α, Ω, ε, P)\}$ of PFSA is said to be a semigroup of PFSA if $T^n T^m = T^{n+m}$, $∀n, m ∈ N$. Thus, a semigroup of PFSA $\{R^n\}$ on a probability space $(Ω, ε, P)$ is said to be measure-preserving and ergodic if $T^n$ is measure-preserving and ergodic $∀n ∈ N$. A straightforward result, which is central to the current paper, follows from Theorem 2.1 for PFSA and is presented as the following corollary.

Corollary 1. Let $\{R^n\}$ be a semigroup of measure-preserving PFSA on a probability space $(Ω, ε, P)$. Then, $\{R^n\}$ is ergodic if and only if the absolute value of every left eigenvector $v_j(n)$ of $T^n$ $∀n $ $∈ $ $N$ and $j = 1, \cdots, M ≤ |Ω|$, is uniformly distributed, i.e.,

$$|v_j(n)| = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

3.2 D-Markov Machines

In the construction of 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 (18) is defined as follows.

Definition 8. A D-Markov machine is a PFSA in the sense of Definition 4 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 $S = \cdots s_{−D} s_{−1} s_0 s_1 \cdots$, where the probability of occurrence of a new symbol depends only on the last consecutive (at most) $D$ symbols, i.e.,

$$P[s_n | \cdots s_{n−D} \cdots s_{n−1}] = P[s_n | s_{n−D} \cdots s_{n−1}] \quad (2)$$

Consequently, for $w ∈ Α^D$ (see Definition 4), the equivalence class $Α^* w$ of all (finite-length) words of suffix $w$, is qualified to be a D-Markov state that is denoted as $w$.

Remark 3. While the algebraic structure of the PFSA in D-Markov machines is deterministic (17)(18), a hidden Markov model (HMM) may have a non-deterministic algebraic structure (16)(17); this difference yields significant computational advantages of D-Markov machines over HMMs at the expense of limited loss of modeling flexibility. Moreover, since HMMs are typically trained by expectation maximization (5), the underlying algorithms might suffer from having a poor local optimum. In addition to its iterative computation cost, HMMs may not be sufficiently robust in terms of convergence even to a locally optimum point. 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 by frequency counting (18), for example.

The PFSA of a D-Markov machine is capable of generating symbol strings. That is, such a generated symbol string
has the form \( \{s_1 s_2 \cdots s_\ell \} \), where \( s_j \in \mathcal{A} \) and \( \ell \) is a positive integer. Both morph function \( \pi \) and state probability transition matrix \( \mathcal{T} \) implicitly support the fact that PFSA satisfies 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 [18].

For construction of the proposed \( D \)-Markov-based PFSA, there are four primary choices as enumerated below:

1. **Alphabet size (|\( \mathcal{A} \)|):** Larger is the alphabet size, the dynamical system is more discretely represented as different symbols, which will require more training data. As such a critical step of PFSA construction is selection of the alphabet \( \mathcal{A} \); while there are several techniques for selection of \( \mathcal{A} \) (e.g., [14][18]), the alphabet size \( \mathcal{A} \) has been chosen to be very small (e.g., \( |\mathcal{A}| = 2 \) or \( 3 \)) in this paper so that the number of states \( |\Omega| \) is also small to limit the required window length \( L \).

2. **Partitioning Method:** While there are many data partitioning techniques, maximum entropy partitioning (MEP) [12][13] and K-means partitioning [29] have been chosen as they are commonly used for symbolization of time series.

3. **Depth (\( D \)) in the \( D \)-Markov machine:** In this paper, \( D = 1 \) has been primarily chosen for PFSA construction to limit the window length \( L \). 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 may need for more training.

4. **Choice of Feature:** The feature needs to be the one that best captures the physical nature (e.g., texture) of the signal and also is not computationally too expensive. In this paper, (absolute sum-normalized) left eigenvectors of the state transition matrix are selected as features.

3.3 **Anomaly Detection in the Standard STSA Setting**

From the perspectives of discrete-time measurements and their discrete-space symbolization, 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 [30], a time series \( \{x_n\} \) is first converted into a symbol string. Then, PFSA 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 partitioning for the nominal block and convert it into a symbol string to construct the nominal PFSA. The emission matrix (and hence the state transition probability matrix) of the PFSA are constructed by frequency counting [18]. 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 \( v^m \) that represents the feature vector at the time epoch \( n \).

4. **Compute** the (scalar-valued) anomaly metric \( \theta(n) \) at the time epoch \( n \) from a string of divergences between the nominal feature vector \( v^0 \) and the current feature vector \( v^m \) and by sliding the block of data \( N \) times as:

\[
\theta(n) \triangleq \frac{1}{N} \sum_{m=n-1}^{n+N-1} d_{KL}(v^0, v^m) \tag{3}
\]

where \( d_{KL}(\cdot, \cdot) \) is the Kullback-Leibler divergence [8], and the number \( N \) of sliding windows serves the sole purpose of data smoothing. The anomaly metric \( \theta(n) \) in Eq. (3) is used in this paper for the standard STSA, which is more robust to outliers and fast fluctuations in the time series than if an individual term is used for computation of \( d_{KL}(v^0, v^m) \).

4 **Technical Approach**

A major issue in the standard STSA-based anomaly detection is that it may require a long time series to construct a stationary state transition probability matrix \( \mathcal{T} \), from which the stationary state probability vector \( v \) is generated to serve as a feature vector. However, many applications do require low-delay detection of anomalous events, for which short-length windows of time series must be used. For example, fatigue damage in critical mechanical structures must be detected as early as possible to avoid a catastrophic failure. This requirement mandates early detection of the damage by using short-length windows for PFSA construction [7].

To address the above issue, the dynamical system is described by an ergodic semigroup of MPTs \( \{\mathcal{T}^n\}, n \in \mathbb{N} \), acting on a probability space \( (\Omega, \mathcal{E}, P) \). It is stated in Subsection [3.1] and is reiterated here that the sample space \( \Omega \) is the (finite) state space \( \Omega \) of the PFSA of a \( D \)-Markov machine, the associated \( \sigma \)-algebra \( \mathcal{E} \) is the power set \( 2^\Omega \), and \( P \) is the probability measure. Based on the concept of MPT-based STSA, if a time-varying PFSA is constructed using short-length windows of time series, then the resulting semigroup of state transition probability matrices \( \{\mathcal{T}^n\} \) would describe non-homogeneous Markov-chain models. Although the underlying stochastic process could be stationary, short-length windows of time series may produce time-varying \( \mathcal{T} \)-matrices; hence, in general, the eigenvalues and eigenvectors of \( \mathcal{T} \) may be time-varying as well. However, in light of Corollary [1] in Subsection [3.1] the left eigenvectors of \( \{\mathcal{T}^n\} \) should have \( n \)-invariant absolute values to reflect measure-invariance and ergodicity of the dynamical system model. Therefore, rather than increasing the length of windows until a constant \( \mathcal{T} \) is obtained, as required by the standard STSA, the proposed MPT-based STSA requires that \( \mathcal{T} \) may only satisfy the property of time-invariant left eigenvectors given by Corollary [1]. Moreover, a metric of variability of left eigenvectors is proposed as a measure of anomaly in the dynami-
A modification of the standard STSA is now proposed based on the variations of eigenvectors. Table 1 lists the major differences, demonstrated in the sequel, between the standard STSA and the proposed MPT-based STSA.

### Table 1: Comparison of standard STSA and MPT-based STSA

<table>
<thead>
<tr>
<th>Standard STSA</th>
<th>MPT-based STSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 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. Generation of homogeneous Markov-chain models</td>
<td>Generation of non-homogeneous Markov-chain models</td>
</tr>
<tr>
<td>3. 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. 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. Less robust to parametric changes in data partitioning and detection system (e.g., alphabet size</td>
<td>More robust to parametric changes in data partitioning and detection system (e.g., alphabet size and Markov depth)</td>
</tr>
</tbody>
</table>

Algorithm 1 and Algorithm 2 implement the theory of proposed MPT-based STSA. Algorithm 1 presents a procedure for identification of the window length $L$ when the measure-preserving and ergodicity conditions are satisfied. In Algorithm 1, the parameter $L$ is kept on increasing until a metric of variations of the left eigenvector, hereafter called ergodicity metric $\rho$, which should be a constant for measure-preserving and ergodic processes, is less than the (user-specified) threshold parameter $\tau_1$. Algorithm 2 presents the steps for detecting anomalous patterns using short-length windows by constructing a D-Markov machine 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. Figure 1 and Figure 2 show the flowcharts of Algorithm 1 and Algorithm 2 respectively.

The threshold parameters $\tau_1$ and $\tau_2$ in Algorithm 1 and Algorithm 2 respectively, are interdependent in the sense that a smaller $\tau_1$ would result in a larger window length $L$ (i.e., increased delay), which would require a smaller $\tau_2$ under a fixed maximum allowable false positive rate (FPR) for similar qualities of decisions on anomaly detection. 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 $\tau_2$.

---

1 Although the theory of MPT-based STSA requires all the left eigenvectors to have time-invariant absolute values, it turns out that one may need to check only one of these left eigenvectors. In particular, it is demonstrated on experimental data in Section 5 that once anomaly occurs in a stochastic process, all the left eigenvectors deviate from having time-invariant absolute values. Apparently, it suffices to look at one of these left eigenvectors in order to check for anomaly. However, depending on the particular application, changes in some of these left eigenvectors could be more effective than the others in quantifying anomaly of the system, in which case the best eigenvector can be chosen based on a held-out training data set.

#### Algorithm 1: MPT-based selection of window length

**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 $\tau_1 > 0$.  
**OUTPUT:** Window length $L$ of the data blocks $\{x_{n+1:n+L}\}$ for construction of a $D$-Markov model of the stochastic process.

1. Choose an initial window size $L$. 
2. do  
3. 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.  
4. Using frequency counting, construct a $D$-Markov machine based on each $s_{n+1:n+L}$ to obtain state transition probability matrices $\{T^n\}$.  
5. 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 $M \leq |\mathcal{A}|$.  
6. $v^i \leftarrow \frac{1}{N} \sum_{n=0}^{N-1} |v^i(n)|$, $i = 1, \ldots, M$, where $|v^i(n)|$ is the vector of the absolute values of the components of $v^i(n)$.  
7. $\theta^i \leftarrow \sum_{n=0}^{N-1} \left|\sum_{i} (|v^i(n)| - v^i)\right|$, $i = 1, \ldots, M$.  
8. $\rho \leftarrow \max_{i \in \{1, \ldots, M\}} |\theta^i| |l_i|$, where $|\theta^i| |l_i| \triangleq \sum_{j=1}^{|\mathcal{A}|} |\theta_j|$.  
9. $L \leftarrow L + \Delta L$.  
10. while $\rho > \tau_1$ then  
11. \hspace{1em} declare the system as normal
12. \hspace{1em} end if

#### Algorithm 2: MPT-based anomaly detection

**INPUT:** Alphabet $\mathcal{A}$, Markov depth $D$, window length $L$, time series $\{x_n\}$, number of sliding windows $N \in \mathbb{N}$, a threshold $\tau_2 > 0$.  
**OUTPUT:** The decision on whether the system is nominal or anomalous.

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.  
2. Using frequency counting, construct a $D$-Markov machine based on each $s_{n+1:n+L}$ to obtain state transition probability matrices $\{T^n\}$.  
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 $M \leq |\mathcal{A}|$.  
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 $v^i(n)$.  
5. $\theta^i \leftarrow \sum_{n=0}^{N-1} \left|\sum_{i} (|v^i(n)| - v^i)\right|$, $i = 1, \ldots, M$.  
6. $\rho \leftarrow \max_{i \in \{1, \ldots, M\}} |\theta^i| |l_i|$, where $|\theta^i| |l_i| \triangleq \sum_{j=1}^{|\mathcal{A}|} |\theta_j|$.  
7. if $\rho > \tau_2$ then  
8. \hspace{1em} declare the system as anomalous
9. \hspace{1em} else  
10. \hspace{2em} declare the system as nominal
11. \hspace{1em} end if
Fig. 1: Flowchart for Algorithm 1

Fig. 2: Flowchart for Algorithm 2

Fig. 3: (a) Fatigue testing apparatus. (b) Details of a test specimen.

is recommended as a topic of future research in Section 6.

5 Experimental Validation

This section evaluates the performance of the MPT-based STSA for early prediction of fatigue damage in a polycrystalline metallic alloy material. In this regard, an ensemble of time series has been generated for both nominal (e.g., undamaged) and anomalous (e.g., damaged) conditions; the objective here is to evaluate the performance of the MPT-based STSA, presented in Algorithms 1 and 2, for online anomaly detection using short-length windows and low-dimensional feature vectors. The performance of the MPT-based STSA is compared with those of a standard STSA (e.g., [18]) and an HMM [5] on the same experimental data. In all three methods (i.e., standard STSA, MPT-based STSA, and HMM), the number of sliding windows is taken as \( N = 30 \) (see Eq. (3), Algorithm 1 in this paper, and Algorithm 2 in [7]).

Seventeen experiments have been conducted on the test apparatus, shown in Figure 3(a), which is built upon a computer-instrumented and computer-controlled fatigue testing machine equipped with ultrasonic sensing and optical microscopy. The test specimens (see Figure 3(b)) are made of polycrystalline aluminum 7075-T6 alloy, where each specimen is 3 mm thick and 50 mm wide with a notch of 1.58 mm × 4.57 mm on one side of the specimen. The notch is made to increase the (local) stress concentration factor that ensures crack initiation and propagation at the notch end.

The test specimens have been subjected to sinusoidal loading on the apparatus under tension-tension stress at a frequency of \( \sim 50 \) Hz. The ultrasonic time series collected during each experiment contains approximately 1,000,000 data points.
points. Since it is not possible by inspection to determine the exact point at which a change in the time series has occurred, an interval is identified (within which this change point is located) to serve as the ground truth, against which efficacy of the test methods are evaluated. A part of the time series before that interval and another part of the time series after that interval are selected such that these two parts can be concatenated to form a single time series with a clearly defined instant of change time; the total length of these two parts of the selected time series is ~ 10,000.

Figure 4 shows the reconstructed ultrasonic signals for several typical specimens, and Figure 5 shows a typical sample of the reconstructed ultrasonic signal with a fatigue failure onset at the approximate stress cycle 4,525. The goal here is to promptly detect such an onset point using a short-length window of the ultrasonic signal. The main challenges in doing so are:

1. Fatigue failure phenomena in polycrystalline alloys are stochastic, which increases the possibility of missed detections and false positives due to large uncertainties of (randomly changing) failure onset points in individual specimens, even under identical loading conditions.
2. Typically the change in a typical signal pattern is very small as seen in the expanded view of Figure 5 over a span of 1,000 cycles at the top right-hand corner. The challenge is to detect such pattern changes and to identify the onset point in near-real time (e.g., within a short window of length, say, ranging from $L = 50$ to $L = 200$).
3. Low-cost ultrasonic sensors have been used for generating and measuring the signals. These sensors are noisy and are very sensitive to the location where they are fixed on the specimen. Therefore, a slight change in their relative locations may produce a significant change in the measured signals. These phenomena would increase the uncertainty of the failure onset point and make its detection even more challenging.

Considering the healthy part of an ultrasonic signal in Figure 5, nearly all the signal components are available at the receiver sensor. The signal energy is nearly constant and the process can be considered as nearly measure-preserving. Moreover, if the signal space is partitioned and the signal time series is converted to a symbol sequence, then every cell is visited many times if the healthy state prevails for a long time period. Therefore, the resulting symbolic dynamics can be considered as measure-preserving and ergodic. In contrast, the transient part of the signal cannot be considered as measure-preserving because some of the signal components are reflected back rather than being available at the receiver, and the received signals are attenuated as a result of the decrease in energy. Consequently, by following Corollary 1 in Subsection 3.1, it is expected that eigenvectors would tend to have time-invariant absolute values in the healthy zone and time-varying absolute values in the transient (anomalous) zone.

The above facts are demonstrated experimentally in Figures 6 and 7 which show the behavior of the absolute values of the eigenvectors of the state transition probability matrix $\mathcal{F}$ over time for different values of window length $L$ of the time series blocks used to construct the state transition probability matrix $\mathcal{F}$, where an alphabet size of $|\mathcal{A}| = 2$ with Markov depth $D = 1$ are used, which leads to the number of states $|Q| = 2$ (see Subsection 3.1). Since $|Q| = 2$ in this case, there are only two eigenvectors, each one of dimension 2. Figure 6 shows the results for the absolute values of the left eigenvectors corresponding to the maximum eigenvalue (i.e., $\lambda = 1$), while Figure 7 shows the results for the absolute values of the eigenvectors corresponding to the minimum eigenvalue. As shown in Figures 6 and 7, the absolute values of the eigenvectors tend to become time-invariant in the healthy phase as the window size $L$ increases, while they are still time-varying on the transient phase. These results are consistent with Corollary 1 that the absolute values of the eigenvectors are time-invariant if the underlying stochastic process is measure-preserving and ergodic. However, the absolute values of the eigenvectors in the healthy phase are not exactly constant over the sample space as suggested by Corollary 1 because the system is not perfectly measure-
preserving as the ultrasonic signal may not be perfectly received without any dissipation. Therefore, the system in the healthy state is considered as nearly measure-preserving. A property which can no longer be maintained for the transient phase where energy dissipation is much more significant. Moreover, it is shown in both Figures [6] and [7] that once the damage evolves starts, both eigenvectors become time-varying and lose the time-invariance property. Therefore, only one of these two eigenvectors is needed for anomaly detection.

Figure [8] and Figure [9] show the Receiver Operating Characteristic (ROC) performance of MPT-based STSA, described in Algorithm 2, compared with the performances of the standard STSA and HMM in terms of the area under the curve (AUC) of ROC plots, with different parameters (i.e., window length \( L \) and alphabet size \( A \)) and partitioning methods. While the results of MPT-based STSA and standard STSA are different for MEP and K-means partitioning in the ROC plots of Figures [8] and [9], respectively, the results for HMM in these two figures are largely similar because data partitioning is not relevant in the (stochastic) algorithm of HMM. All eight plates in Figures [8] and [9] exhibit that the performance of MPT-based STSA is consistently superior to those of a standard STSA [18] and a hidden Markov model (HMM) [5,6,7] for anomaly detection; the results show consistent superiority of MPT-based STSA in terms of detection accuracy. Whereas standard STSA and HMM require learning a nominal model, with which a new generated model is compared, this is not needed in MPT-based STSA. This property of MPT-based STSA enhances robustness of decisions on anomaly detection to changes in the nominal phase.

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. Pareto optimization for selection of the threshold parameters \( \tau_1 \) and \( \tau_2 \) in Algorithms [7] and [2], respectively: This will also involve rigorous statistical analysis (e.g., \[33\])
2. Comparative evaluation of MPT-based STSA with other techniques of anomaly detection: Examples of potential candidates are neural network-based forecasting \[44\] and dynamic mode decomposition (DMD) \[35\].
3. Investigation of sufficient conditions for commutativity of MPTs: In this case, the commutator norm can be used as a measure of evolving anomalies, which would yield a significant reduction in CPU computation time.
4. Experimental validation of MPT-based STSA in diverse applications: These applications should demonstrate high performance and robustness of MPT-based STSA for real-time monitoring and active control.

### Acknowledgements

The work reported in this paper has been supported in part by U.S. Air Force Office of Scientific Research (AFOSR) under Grant No. FA9550-15-1-0400 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.

---

\[3\] The Receiver Operating Characteristic (ROC) curves have been used here for assessing the detection performance by varying the parameter \( \tau \) from \(-\infty \) to \( \infty \) \[31\], where each point in a ROC curve corresponds to a specific value of the threshold \( \tau \). It is noted that \( \tau \) can be determined from the ROC curve by specifying a maximum allowable false positive rate (FPR), which may depend on the application. 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.

\[4\] For fair comparison, the number of states for both STSA methods and HMM in each experiment is the same. That is, if an alphabet \( A \) is used with a Markov depth \( D \) in STSA, then \( |A|^D \) hidden states are used in HMM.

\[5\] The results have been computed on a DELL PRECISION T3400 with an Intel(R) Core(TM)2 Quad CPU Q9550 at 2.83 GHz and 8 GB RAM, running under Windows 7.

\[6\] The experimental validation has been conducted with different detection parameters, data partitioning methods, and window lengths of the observed time series. Performance of the proposed MPT-based STSA has been compared with those of a standard STSA \[18\] and a hidden Markov model (HMM) \[5,6,7\] for anomaly detection; the results show consistent superiority of MPT-based STSA in terms of detection accuracy. Whereas standard STSA and HMM require learning a nominal model, with which a new generated model is compared, this is not needed in MPT-based STSA. This property of MPT-based STSA enhances robustness of decisions on anomaly detection to changes in the nominal phase.
Fig. 6: Convergence of eigenvectors, corresponding to the maximum eigenvalue ($\lambda = 1$) of an ultrasonic signal to a constant in the healthy phase, and to be oscillatory in the damage phase ($|\lambda| = 2, D = 1$, and different values of window size $L$).

Fig. 7: Convergence of eigenvectors, corresponding to the minimum eigenvalue to constants in the healthy and completely broken phases, and to be oscillatory in the damage phase ($|\lambda| = 2, D = 1$, and different values of window size $L$).

<table>
<thead>
<tr>
<th>Parameters &amp; partitioning method</th>
<th>Standard STSA</th>
<th>MPT-based STSA</th>
<th>HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\lambda</td>
<td>= 2, D = 1, L = 50, MEP$</td>
<td>2.133443e-04</td>
</tr>
<tr>
<td>$</td>
<td>\lambda</td>
<td>= 2, D = 1, L = 100, MEP$</td>
<td>3.865693e-04</td>
</tr>
<tr>
<td>$</td>
<td>\lambda</td>
<td>= 2, D = 1, L = 200, MEP$</td>
<td>9.031112e-04</td>
</tr>
<tr>
<td>$</td>
<td>\lambda</td>
<td>= 3, D = 1, L = 200, MEP$</td>
<td>1.733018e-03</td>
</tr>
<tr>
<td>$</td>
<td>\lambda</td>
<td>= 2, D = 1, L = 50, K-means$</td>
<td>1.053167e-03</td>
</tr>
<tr>
<td>$</td>
<td>\lambda</td>
<td>= 2, D = 1, L = 100, K-means$</td>
<td>1.339111e-03</td>
</tr>
<tr>
<td>$</td>
<td>\lambda</td>
<td>= 2, D = 1, L = 200, K-means$</td>
<td>2.177394e-03</td>
</tr>
<tr>
<td>$</td>
<td>\lambda</td>
<td>= 3, D = 1, L = 200, K-means$</td>
<td>3.047629e-03</td>
</tr>
</tbody>
</table>

Table 2: Statistics of CPU execution time (sec) per window for online fatigue damage detection

References
AUC: Standard STSA = 0.60445, HMM = 0.82525, MPT-based STSA = 0.88399

AUC: Standard STSA = 0.60499, HMM = 0.93907, MPT-based STSA = 0.98909

AUC: Standard STSA = 0.90028, HMM = 0.87539, MPT-based STSA = 0.99893

AUC: Standard STSA = 0.66925, HMM = 0.90078, MPT-based STSA = 0.95129

Fig. 8: ROC performance of the MPT-based STSA (using eigenvectors corresponding to the minimum eigenvalue) and the standard STSA, with MEP partitioning, and and HMM [6,7] for fatigue damage detection with (a) |A| = 2, D = 1, and L = 50, (b) |A| = 2, D = 1, and L = 100, (c) |A| = 2, D = 1, and L = 200, (d) |A| = 3, D = 1, and L = 200.


[23] Azer, A., Siriwardane, S., Mikkelsen, O., and Lan-
Fig. 9: ROC performance of the MPT-based STSA (using eigenvectors corresponding to the minimum eigenvalue) and the standard STSA, with K-means partitioning, and HMM [6, 7] for fatigue damage detection with (a) $|A| = 2$, $D = 1$, and $L = 50$, (b) $|A| = 2$, $D = 1$, and $L = 100$, (c) $|A| = 2$, $D = 1$, and $L = 200$, (d) $|A| = 3$, $D = 1$, and $L = 200$.