Mechanical Systems and Signal Processing 23 (2009) 358–371

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/jnlabr/ymssp

An information-theoretic measure for anomaly detection in complex dynamical systems ☆

Abhishek Srivastav, Asok Ray *, Shalabh Gupta

The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 11 April 2007
Received in revised form
30 March 2008
Accepted 1 April 2008
Available online 2 June 2008

Keywords:
Anomaly detection
Equilibrium thermodynamics
Statistical mechanics
Fatigue damage
Duffing oscillator

ABSTRACT

This paper presents information-theoretic analysis of time-series data to detect slowly evolving anomalies (i.e., deviations from a nominal operating condition) in dynamical systems. A measure for anomaly detection is formulated based on the concepts derived from *information theory* and *statistical thermodynamics*. The underlying algorithm is first tested on a low-dimensional complex dynamical system with a known structure—the Duffing oscillator with slowly changing dissipation. Then, the anomaly detection tool is experimentally validated on test specimens of 7075-T6 aluminum alloy under cyclic loading. The results are presented for both cases and the efficacy of the proposed method is thus demonstrated for systems of known and unknown structures.

1. Introduction

Anomaly in a dynamical system is defined as a deviation of the system performance from the expected or the nominal behavior. Usually resulting from slow parametric or non-parametric changes within the system, growth of anomalies often leads to degraded performance and eventually premature end of service life. Anomaly detection is essential for sustaining order and normalcy in human-engineered complex systems.

Anomaly detection and fault estimation in engineering systems can be broadly classified into two categories—model based and dynamic data-driven. In the model-based category, observer-based techniques are commonly used, where certain *residuals* or diagnostic signals are generated for use in optimal or adaptive threshold functions to detect the presence of faults. Residuals are generated by estimating system's measured variables using a deterministic observer or a stochastic filter [1]. These observers are often designed based on a linear model or a Jacobian linearization of the system model at the operating points. These linear or linearized models are reasonable approximations of the complex system when operating under the nominal condition; however, in case of evolving anomalies system nonlinearities may become too large to be ignored or approximated. Furthermore, it is very difficult to model anomalous system behavior because such anomalies (e.g., incipient faults) are usually unknown and are too complex to model. Furthermore, the fault might be entering the system in a more complex manner than as an additive signal. These issues have motivated the study of anomaly detection in dynamical systems using a data-driven approach, where the dependence on a physics-based system model is reduced. Various data-driven methods have been reported in literature for fault diagnosis and failure prognosis.

☆ This work has been supported in part by the Army Research Laboratory and the Army Research Office under Grant no. W911NF-07-1-0376, by the Office of Naval Research under Grant no. N00014-08-1-380, and by NASA under Cooperative Agreement No. NNX07AK49A.

* Corresponding author. Tel.: +1 814 865 6377; fax: +1 814 863 4848.
E-mail addresses: axs964@psu.edu (A. Srivastav), axr2@psu.edu (A. Ray), szg107@psu.edu (S. Gupta).

0888-3270/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ymssp.2008.04.007

Examples are linear classification techniques (e.g., support vector machines (SVM) [2] and principal component analysis (PCA) [3]), methods based on nonlinear dynamical systems [4–7], neural networks [8], and statistical methods [9]. This paper follows a statistical data-driven approach where the dynamical system is treated as a black-box spewing time-series data that are used to extract pertinent information about system characteristics and monitor its behavior with evolution in time. The knowledge of a physics-based model, if available, will supplement the information generated from the measured time-series data.

Recent literature [10,11] has reported a statistical pattern identification technique for anomaly detection, based on symbolic time-series analysis (STSA) [12]. A comparative evaluation of this novel analytical method demonstrates its superior performance relative to other existing pattern recognition tools, such as PCA and artificial neural network (ANN) in terms of early detection of small changes in dynamical systems [13,14]. A new technique of pattern identification for tracking anomaly evolution, based on the concepts of statistical thermodynamics [15] and information theory [16], has been presented by the authors in recent conferences [17,18]. This paper consolidates and enhances this concept to build a rigorous theory based on thermodynamic formalism of complex systems for anomaly detection.

The theoretical concepts, presented in this paper, are first applied to analysis of a low-dimensional nonlinear system—the Duffing oscillator with slowly varying dissipation parameter [10,19]. This approach serves as a test case for proof of the concept, where the system structure is known and the expected behavior is understood and has been well-studied in available literature [20]. The validated concept is then applied to monitor fatigue damage in polycrystalline alloys (e.g., 7075-T6 aluminum alloy) by experimentation on a laboratory apparatus [14]. It is shown that the proposed method can be effectively used to monitor and detect incipient faults, eventually manifesting themselves as phase transitions in the thermodynamic sense.

The rest of the paper is organized as follows. Section 2 presents the details of the analytical formulation. Theoretical formulation of the proposed methodology is tested by simulation on a Duffing oscillator in Section 3. The method is experimentally validated in Section 4 for anomaly detection on a laboratory apparatus to detect fatigue failure. The paper is summarized and concluded in Section 5 with suggestions for future work. Necessary background materials that have been reported in detail in earlier publications are succinctly reported in the appendices for the sake of completeness of the work presented in the main text.

2. Methodology

This section presents the methodology formulated in this paper to analyze (possibly) anomalous behavior of complex dynamical systems based on the concept of thermodynamic formalism [21,22] as explained below.

Classical equilibrium thermodynamics deals with the study of *macro-level* description of evolution as certain parameters change *quasi-statically* relative to the *micro-level* response time of the dynamical system. Therefore, the macro-properties of the quasi-static system are described while being oblivious to the underlying micro-motion. The bridge between the micro- and macro-properties is provided by statistical thermodynamics where a statistical approach is used to study the micro-motion based on the observed macro-values. In essence, a statistical thermodynamic approach is suitable for studying any system where it is possible to make a distinction between micro-motion and macro-behavior.

In case of human-engineered dynamical systems, *fast* dynamics of the system can serve as the micro-motion, while the *slow* dynamics that affect the macro-behavior could be, for example, due to drifts in critical parameters or may result from gradual changes in the environment. Usually, in these cases, the macro-behavior and (possibly) the micro-motion can be observed from direct measurements (e.g., sensing devices) or analytically derived from other direct measurements. This is unlike the case of classical statistical thermodynamics of gases, where only the macro-variables can be observed and the micro-motion is not directly measurable; the statistics of micro-motion are estimated from the measured macro-values. Therefore, in human-engineered systems, both the problem of describing the macro-behavior based on micro-motion (forward analysis) and the problem of estimating the micro-motion statistics using observed macro-values (inverse analysis) are well-defined. This forms the backbone of the concept presented in this paper, where a combination of both the forward and the inverse analysis is used to construct a measure of net information gain (or entropy loss) to monitor anomaly evolution in a dynamical system.

The key idea of the work reported in this paper is to utilize the concepts from statistical thermodynamics and information theory for anomaly detection by data-driven time-series analysis of observables, generated from one or more sensors, that does not require prior knowledge of the system model. It should be noted that the major motivation for thermodynamic formalism is to deal with complexity of the nonlinear system under consideration, not necessarily high dimensionality of the state space.

Entropy (or information) is a pivotal concept in statistical thermodynamics and dynamical systems for quantifying the degree of uncertainty. Recent literature has reported applications of information-theoretic techniques for analysis of nonlinear dynamical systems. Examples are mutual information to study synchronization among coupled systems [23], transfer entropy to detect nonlinearities due to evolving anomalies [24], and direction of information flow from one subsystem to another [25].

Information is generated as the system dynamics evolve over time. Given a reference probability density, information gain (or entropy loss) can be quantified by using a measure of relative information gain (e.g., Kullback–Liebler [16]). In this

paper, the reference probability density is estimated from the current macroscopic observations, such as energy, instead of a fixed nominal density. The evolution of (possible) anomalies could be monitored from statistics of the system trajectory in the phase space [26] or in a transformed space (e.g., wavelet space [10,11]).

Sensor data can be generated from a self-excited system, such as vibration data from a mechanical structure, or in response to external stimuli such as ultrasonic excitation. In both cases, the task of anomaly detection in complex systems is identified as a two-time-scale problem that separates micro-motion and macro-behavior. This concept of a two-time-scale approach has been reported in literature [10,27] and the references cited therein, to decouple the *slow* parametric drifts (anomalies) present in system from the relatively stationary fast dynamics of the system. A similar approach is followed here in this paper and the following assumptions are made.

- On the time scale of micro-motion, called the *fast scale*, the system behavior is quasi-stationary.
- Relative to the fast scale of micro-motion, parametric or non-parametric changes in the system evolve on a much (e.g., several orders of magnitude) slower time scale, called the *slow scale*.

The choice of the time scales in the above assumptions is dependent on the dynamics of the system under consideration and the rate of change of system behavior (i.e., possible anomalies). For example, an anomaly may evolve in the slow scale of hours due to changes in the critical parameter(s) of a mechanical system vibrating on the fast scale of milliseconds. Following conclusions are drawn based on the above two assumptions.

- A long time span in the fast scale is a short (e.g., several orders of magnitude smaller) duration in the slow scale. The time span in the fast scale, over which the statistics of micro-motion remain stationary, is called an *epoch* in the slow scale.
- Change(s) in macro-properties that are indicative of anomaly, evolve in the slow time scale. Thus, an anomaly remains constant over an *epoch*.

2.1. Space partitioning and energy levels

To extract the statistics of micro-motion of a given dynamical system, the phase space of the system is partitioned into cells of appropriate size. The region of interest is usually contained within a compact (i.e., closed and bounded) region of the phase space, which is identified and partitioned. The feature of interest here is not the exact locations in the phase space, visited by the system in *fast time*, but the relative frequency with which different regions of the phase space are visited over a given period of time (*epoch*). This approach is not only a tractable choice but also it is believed to preserve the basic characteristics of the system dynamics with an appropriate choice of the partition [28].

Several partitioning techniques have been reported in literature for coarse graining of the state space [12,26] primarily based on symbolic false neighbors. These techniques may become cumbersome and computation-intensive if the dimension of the state space is large. Moreover, if the time-series data are noise-corrupted, then the symbolic false neighbors would rapidly grow in number and require a large symbol alphabet to capture the pertinent information on the system dynamics. This paper has adopted a *wavelet-based* partitioning where the wavelet transform [29] largely alleviates the above shortcomings and is particularly effective with noisy data from high-dimensional dynamical systems. The motivation behind using a wavelet-based approach is the ability of the wavelet transform to perform flexible localization both in time and frequency. This ability is relevant in the area of anomaly detection. The choice of wavelets is dependent on the system under consideration and the frequency range of interest in the observed data sequence. A detailed formulation and evaluation of the *wavelet-based* partitioning method adopted in this paper is presented in [11] (Appendix B), while a literature review on various applications of wavelet transform for fault diagnostics can be found in [30].

The wavelet-transformed data are partitioned by using the *maximum entropy principle* [11] such that the regions rich in information are partitioned finer and those with sparse information are partitioned coarser. Maximum entropy is achieved by a partition that produces a uniform probability of all partitions (cells). Partition once obtained is kept fixed for further calculations. Fig. 1 shows an illustrative example, where the partitioning has been done to create 10 cells (partitions), the size of the cells (width in this case) is smaller for regions with higher density of data points. This ensures an unbiased partition such that each cell gets equal number of visits to begin with. Sensor time-series data are used to compute the estimated statistical distribution of cells visited by the system trajectory in an *epoch* [10,11]. Therefore for m cells the vector $\hat{\mathbf{p}}^k = [\hat{p}_1^k, \ldots, \hat{p}_m^k]$ at an epoch k, as the observed probability distribution, is computed by a relative frequency count with a stopping rule [31] (Appendix B.3). Therefore each $\hat{p}_j^k \in [0,1]$ and $\sum_{j=1}^{m} \hat{p}_j^k = 1$.

Each cell j of the partitioned space is then identified with a value M_j of a macro-variable M such that so long as the system trajectory stays within the cell j, the averaged characteristics of the system are given by M_j. Therefore, M is treated as a random variable that takes the value M_j with probability p_j, i.e., the probability of being in cell j is p_j. An intuitive choice of the macro-variable M is energy, which is chosen for the formulation presented in this paper. Obviously, there could be

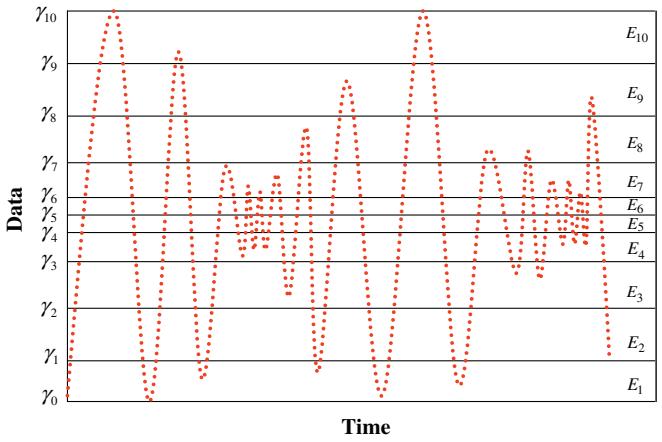

Fig. 1. An example of maximum entropy partitioning.

other possible choices of the macro-variable M to define the equivalence classes; finding the criteria for an appropriate selection is a topic of current research.

Given a data sequence $Y = \{y_i\}_{i=1}^N$ at the nominal condition, let $\Gamma \triangleq \{\gamma_0, \gamma_1, \ldots, \gamma_m\}$, where $y_{\min} = \gamma_0 < \gamma_j < \gamma_{j+1} < \gamma_m = y_{\max}$; $j = 1, 2, \ldots (m-2)$ be the maximum entropy partition [11] of the space $\Omega \triangleq [y_{\min}, y_{\max}]$ that creates m-cells as seen in Fig. 1. Then, the energy E_j associated with the cell j (corresponding to the region $[\gamma_{j-1}, \gamma_j]$) is chosen to be a function φ of the cell boundary points as

$$E_j \triangleq \varphi(\gamma_{j-1}, \gamma_j) = \left(\frac{\gamma_{j-1}^2 + \gamma_j^2}{2} \right) \tag{1}$$

If the data points are uniformly distributed within a cell for every cell of the partition, the error between the actual average energy ($\equiv (1/N)\sum_{i=1}^N y_i^2$) of the signal and its estimate, obtained as the expectation $\mathbf{E} \triangleq \sum \hat{p}_j E_j$, is minimized for $E_j, j = 1, 2, \ldots, m$, defined in Eq. (1). This is because E_j is the arithmetic mean of energies $(y_{i_j})^2$ of data points in the cell j and

$$\Delta \triangleq \left| \sum \hat{p}_j E_j - \frac{1}{N} \sum_i^N y_i^2 \right| = \left| \frac{1}{N} \sum_j^m n_j E_j - \frac{1}{N} \sum_j^m \sum_{i_j} (y_{i_j})^2 \right| \leqslant \frac{1}{N} \sum_j^m \left| \sum_{i_j} [E_j - (y_{i_j})^2] \right| \tag{2}$$

where N is the total number of data points collected in that specific epoch; n_j is the number of times the cell j is visited; and i_j's are the indices of the n_j data points belonging to energy level j. That is, $\{y_{i_j}\}_{j=1}^{n_j}$ is the set of data points belonging to cell j and is a subsequence of the sequence $\{y_i\}_{i=1}^N$.

If the distribution within the cells is non-uniform, then it follows from Eq. (2) that

$$\begin{aligned} \Delta &\leqslant \frac{1}{N} \sum_j^m \sum_{i_j} |E_j - (y_{i_j})^2| \leqslant \frac{1}{N} \sum_j^m n_j \max_{i_j} |E_j - (y_{i_j})^2| \\ &= \sum_j^m \hat{p}_j \frac{(\gamma_j^2 - \gamma_{j-1}^2)}{2} \leqslant \max_j \frac{(\gamma_j^2 - \gamma_{j-1}^2)}{2} = C \end{aligned} \tag{3}$$

The error Δ in the estimation of the expected energy $\mathbf{E}$ is bounded above by a constant that is dependent on the size of the largest cell in the partition. The estimation error Δ is usually reduced by increasing the number of cells. However, there are upper and lower bounds on the number of cells. The upper bound is dependent on the number of data points, which is dictated by stationarity of the estimated probability vector $\hat{\mathbf{p}}^k$ (Appendix B.3), and the lower bound is a function of system dynamics (Appendix B.2).

Therefore, the above construction realizes each partition $[\gamma_{j-1}, \gamma_j]$ as an energy level j such that a data point belonging into this region has an average energy E_j. From a statistical thermodynamic point of view, a point in the phase space corresponds to a micro-state of the system while an energy level j is an equivalence class of micro-states characterized by the same energy value E_j. Therefore, during an *epoch* k the system trajectory can be said to visit several micro-states (or points in phase space) of the system. Grouping the micro-states into equivalence classes E_j's based on energy reveals the distribution among the various energy levels E_j's. In the forward calculation of estimating macro-values based on micro-motion statistics $\hat{\mathbf{p}} = [\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_m]$, the macro-value $\mathbf{E}$ is computed as the expectation $\sum_{j=1}^m \hat{p}_j E_j$, this is used as the observed macro-value for analysis in the following sections.

2.2. Generalized canonical distribution

Given an observed value of macro-observation, the principles of *generalized canonical distributions* and *entropy maximization* are applied to estimate the probability distribution of the energy levels (i.e., micro-motion statistics) [21]. While a general approach for more than one macro-variable is provided in Appendix A, the probability distribution, for the choice of energy as the macro-variable, is presented below.

Let E be a random variable that takes a value E_j in level j. Let p_j be the unknown probability of this equivalence class j of micro-states. The task is to estimate the probabilities p_j's given a macro-value $\mathbf{E}$ for the system (from previous section). Therefore,

$$\mathbf{E} = \sum_{j=1}^{m} p_j E_j, \qquad \sum_{j=1}^{m} p_j = 1 \quad \text{and} \quad p_j \in [0,1] \tag{4}$$

Shannon entropy is defined as

$$S(\mathbf{p}) = -\sum_{j=1}^{m} p_j \ln p_j \tag{5}$$

where $S(\mathbf{p})$ takes a maximum value ($\ln m$) for uniform distribution (i.e., $p_j = 1/m\ \forall j$) as representation of maximum disorder. An *unbiased guess* of probabilities p_j's of the macro-states that must satisfy the constraints in Eq. (4) is computed using the method of Lagrange multipliers. The idea is to maximize the entropy $S(\mathbf{p})$ to ensure an unbiased guess while the constraints given by Eq. (4) are satisfied as delineated in the following pair of equations:

$$\sum_{j=1}^{m} E_j \delta p_j = 0, \quad \sum_{j=1}^{m} \delta p_j = 0 \tag{6}$$

Then, using the method of Lagrange multipliers, the following equation is derived from Eqs. (4) and (5) as

$$\sum_{j=1}^{m} (\ln p_j - \Psi + \beta E_j)\delta p_j = 0 \tag{7}$$

where Ψ and β are the Lagrange multipliers. Since δp_j's are arbitrary, each of the terms in the summation in Eq. (7) must vanish. Therefore, the estimated probabilities P_j's, called the *canonical distribution*, are of the form

$$P_j = \exp(\Psi - \beta E_j) \tag{8}$$

In Eq. (8), P_j's are also known as *escort probabilities* [21].

Given a macro-observation $\mathbf{E}$ and the values $E_j,\ j = 1, 2, \ldots, m$, of the energy levels, the Lagrange multipliers Ψ and β are calculated by solving the following simultaneous set of non-linear equations that are obtained by combining Eq. (4) with Eq. (8):

$$\mathbf{E} = \sum_{j=1}^{m} [\exp(\Psi - \beta E_j)] E_j \tag{9}$$

$$1 = \sum_{j=1}^{m} \exp(\Psi - \beta E_j) \tag{10}$$

Having evaluated Ψ and β, the canonical distribution $\mathbf{P} = [P_1, \ldots, P_m]$ is obtained from Eq. (8). It is noteworthy that Ψ is the negative of the logarithm of the partition function Z and β proportional to the inverse of absolute temperature in the thermodynamic sense [21] (see also Appendix A).

2.3. Net information gain and anomaly measure

In the fields of communication theory and data compression, Shannon information is defined as the negative of Shannon entropy, defined in Eq. (5)

$$I(\mathbf{p}) = -S(\mathbf{p}) = \sum_{j=1}^{m} p_j \ln p_j \tag{11}$$

has been used as the measure of information content in the probability distribution $\mathbf{p}$, which attains the maximum value of zero at $p_j = \delta_{ij}$ for some i (note: $0 \ln 0 \triangleq 0$). In essence, $I(\mathbf{p})$ measures the amount of certainty or order present in the dynamical system that generates the distribution $\mathbf{p}$.

Kullback–Leibler information measure of the probability $\mathbf{p}^k$ at an epoch τ_k relative to the distribution $\mathbf{p}^0$ at the reference epoch τ_0 is defined as

$$K(\mathbf{p}^k, \mathbf{p}^0) = \sum_{j=1}^{m} p_j^k \ln \frac{p_j^k}{p_j^0} \tag{12}$$

and quantifies the information gain obtained by comparison of the distribution $\mathbf{p}^k$ with the reference distribution $\mathbf{p}^0$. So, while the Shannon information is seen as the measure of *absolute* information, Kullback–Leibler information gain is a relative measure from a given reference. The choice of $\mathbf{p}^0$ is arbitrary, and having $\mathbf{p}^0$ equal to the uniform distribution, Kullback information gain degenerates to

$$K(\mathbf{p}, \mathbf{p}^0) = I(\mathbf{p}) + \ln m \tag{13}$$

Now, we define a new measure $\hat{I}(\mathbf{p})$ of net information contained in a probability distribution for a given value of a macroscopic parameter. For an observed value of energy $\mathbf{E}$ contained in the time series, we have two distributions $\hat{\mathbf{p}}$ and $\mathbf{P}$ as derived in the previous sections. While the vector $\hat{\mathbf{p}}$ describes observed distribution of energy among the various energy levels, $\mathbf{P}$ is the estimated canonical distribution that maximizes the entropy for the same value $\mathbf{E}$ of energy. In this context, a measure of net information gain $\hat{I}(\mathbf{p})$ is defined as

$$\hat{I}(\mathbf{p}) = I(\mathbf{p}) - \min_{\mathbf{p}} I(\mathbf{p})\Big|_{\mathbf{E}} \tag{14}$$

It follows from Section 2.2 and Eq. (14) that $\hat{I}(\mathbf{p}) = I(\mathbf{p}) - I(\mathbf{P})$. Therefore,

$$\hat{I}(\hat{\mathbf{p}}) = \sum_{j=1}^{m} \hat{p}_j \ln \hat{p}_j - \sum_{j=1}^{m} P_j \ln P_j \tag{15}$$

It is argued that since the knowledge of a macroscopic observable is a source of information in itself, the information contained in the probability distribution $\hat{\mathbf{p}}$ of energy levels, which achieves a macro-value, is bounded from below. Thus, given a macro-value, the amount of new information available in the observed distribution can be evaluated by using Eq. (14). Since $\hat{I}(\mathbf{p})$ is computed by removing the minimum information possible for a given value $\mathbf{E}$ of energy, its sensitivity to distinguish between systems exhibiting similar macroscopic behavior but differing in micro-motion characteristics is conjectured to be higher. This is due to the fact that $\hat{I}(\mathbf{p})$ tracks the variations above the minimum threshold for that macro-value $\mathbf{E}$. This can provide useful insight on the behavior of a system as it evolves in time.

The following observations are made on the information measure $\hat{I}(\mathbf{p})$ from Eq. (14):

(i) $\hat{I}(\mathbf{p}) \geqslant 0$ for all $\mathbf{p}$.
(ii) $\hat{I}(\mathbf{p}) = \hat{I}(p_1, \ldots, p_m)$, i.e., for a given $\mathbf{E}$, the information measure is a function of the probabilities only.
(iii) $\hat{I}(1/m, \ldots, 1/m) = 0 \leqslant \hat{I}(\mathbf{p})$.
(iv) If all energy states E_j are identical, i.e., $E_j = E$, for $j = 1, 2, \ldots, m$, then $\hat{I}(\mathbf{p})$ degenerates to the special case of Kullback–Leibler information gain given by Eq. (13) and $\hat{I}(\mathbf{p}) = I(\mathbf{p}) + \ln(m)$ because $\min I(\mathbf{p}) = -\ln(m)$ and in this case the macroscopic energy of the system is E $\forall \mathbf{p}$.
(v) For computing the net information gain $\hat{I}(\mathbf{p})$ the choice of the reference distribution is well-defined.

Remark 1. Properties (ii) and (iii) are the first two of the four Khinchin axioms [21].

Given a data sequence $\{y_i^k\}_{i=1}^N$ from a dynamical system for an epoch τ_k we have the estimated probability distribution $\mathbf{p}^k$ and the corresponding canonical distribution $\mathbf{P}^k$, obtained by minimizing $I(\mathbf{p}^k)$ for observed value $\mathbf{E}$ of energy based on formulations in previous sections. Then, the net information gain $\hat{I}^k$ at an epoch τ_k can be obtained as

$$\hat{I}^k = \sum_{j=1}^{m} p_j^k \ln p_j^k - \sum_{j=1}^{m} P_j^k \ln P_j^k \tag{16}$$

Now, we define a cumulative measure μ^k as accumulation of net generated information up to the epoch τ_k as

$$\mu^k = \sum_{i=1}^{k} \hat{I}^i \tag{17}$$

Therefore, while the cumulative measure μ^k provides an estimate of total information accrued till the epoch τ_k, the net information gain $\hat{I}^k$ is an estimate of instantaneous change in the dynamical system at epoch τ_k. Both $\hat{I}^k$ and μ^k have been used for anomaly detection in the two applications presented in the following sections.

3. The Duffing oscillator

This section presents the application of the above developed concept for anomaly detection in a nonlinear active electronic system characterized by the Duffing equation [19,20]. This system will serve as a simulation test-bed to validate and demonstrate potential detection capabilities of the information measure, derived in the previous section. A Duffing oscillator is described by the following second order non-autonomous system.

$$\frac{d^2x(t)}{dt^2} + \nu(\tau)\frac{dx(t)}{dt} + \kappa x(t) + \lambda x^3(t) = A\cos\Omega t \tag{18}$$

where the fast-scale time is represented by t; the nonlinearity is represented by the cubic polynomial $\kappa x(t) + \lambda x^3(t)$; and the quasi-static dissipation parameter ν is allowed to vary in the slow-scale time τ. The system is excited by an exogenous input of magnitude A and frequency Ω. This equation has been reported as the underlying model for many engineering applications such as electronic circuits, structural vibrations, and chemical bonds [32]. The system shows a characteristic change in dynamics with changes in the dissipation parameter ν. This is analogous to what is observed in many engineering systems where a slow change of a system parameter (e.g., increase in friction coefficient or damping constant) results in system degradation taking the system gradually from one mode of behavior (nominal) to a characteristically different mode (i.e., anomalous behavior) in the slow-scale time. The task of anomaly detection points to two different but inter-related problems of practical significance: (i) *diagnosis*, i.e., to detect this modal change of behavior, and (ii) *prognosis*, i.e., to recognize and quantify precursors leading to anomalous modes with the objectives of failure mitigation and performance enhancement.

Parameters for the forcing function in this simulation experiment were set at $A = 22$ and $\Omega = 5$ rad/s. The parameters κ and λ were chosen to be unity while ν was varied from the nominal value 0.02 to the final value 0.40, beyond which no appreciable change in the characteristic performance was observed. To analyze this system following the presented methodology, the position data $x(t)$ for a fixed value of $\nu(\tau_k)$ was collected over an epoch τ_k equivalent to 12 s of operation. The time-series data are first processed by wavelet transform using a Mexican hat basis function 'mexh' in MATLAB®. The scale coefficients thus obtained are partitioned into eight cells. A brief discussion on the choice of scales and wavelet basis functions can be found in Appendix B.1. The value of ν is held constant for each epoch in the slow scale (consisting of 12 s) and the (fast-scale) time-series data are sampled at 1 kHz.

The four phase plots of $x(t)$ versus v $(= dx/dt)$ in the top half of Fig. 2 exhibit evolution of the characteristic behavior of the dynamical system for four different values of dissipation constant, $\nu = 0.02$, 0.20, 0.320 and 0.325. For $\nu \leqslant 0.320$ the system shows one mode of dynamics—with phase-space trajectory showing the formation of two loops. The system undergoes a catastrophic change of characteristic dynamics for $\nu > 0.320$ when the stationary phase trajectory transforms itself into a single loop closed curve. As ν is increased from 0.02 towards the transition point at $\nu = 0.320$ the system phase plots show only a slight variation visually but the dynamical system is gradually organizing itself towards a drastic change. These variations are contained the probability histograms of $\hat{\mathbf{p}}$ (blue), while $\mathbf{P}$ (yellow) is the corresponding canonical distribution, both are shown as bar charts in Fig. 2.

The plot of net information gain $\hat{I}(\hat{\mathbf{p}})$ versus ν is shown in Fig. 3(i). Information revealing small changes leading to catastrophic change is encoded in the probability density $\hat{\mathbf{p}}$, the information gain measure $\hat{I}$ extracts this and computes the

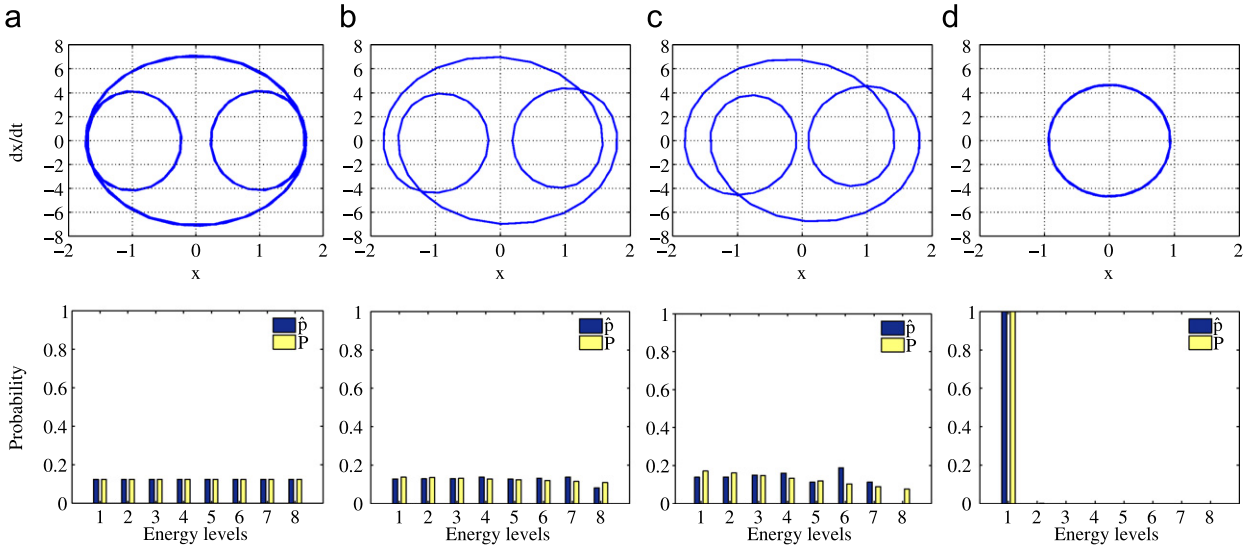

Fig. 2. Phase plot and probability histograms for four values of dissipation parameter ν. (a) $\nu = 0.02$, (b) $\nu = 0.2$, (c) $\nu = 0.320$, (d) $\nu = 0.325$.

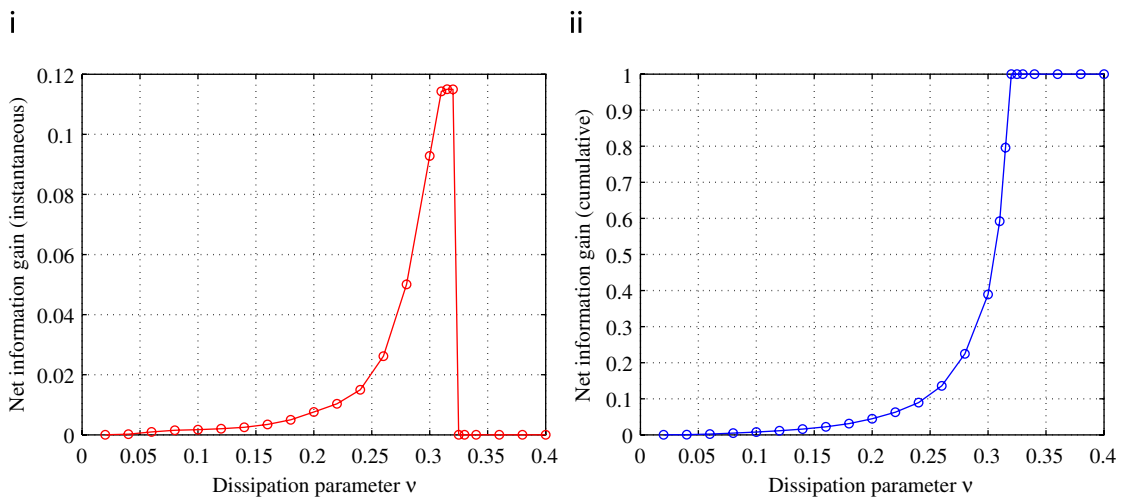

Fig. 3. Net information gain (i) instantaneous $\hat{I}(\hat{\mathbf{p}})$ and (ii) cumulative μ.

net gain of instantaneous information. In Fig. 3(i), $\hat{I}(\hat{\mathbf{p}})$ initially shows a slow rise from $v = 0.02$ to 0.2 after which there is rapid growth leading to the transition point at $v = 0.320$. The plot then shows a discontinuity in $\hat{I}$ at this point and this event can be seen as a phase transition of zeroth order in the thermodynamic sense. To investigate this sudden transition of behavior better, v values were more closely spaced near the transition point (a step of 0.005 instead of the usual 0.02).

A higher value of the information gain $\hat{I}$ implies a higher degree of organization or bias for the given macro-behavior (see Section 2.3). The rationale is that when the dissipation parameter v is slowly increasing towards the transition point at $v = 0.320$, more information is available about the impending catastrophic change at each slow time epoch. Therefore, with increasing values of v, an increasing degree of bias or organization is correctly captured by the plot of $\hat{I}$ versus v. The system becomes stable in the new phase for $v > 0.320$ and no significant change is seen in the phase plots. A similar trend is shown by $\hat{I}$ for the new phase and the $\hat{I}$ values oscillate around a very low value. This can also be read from the histograms of $\hat{\mathbf{p}}$ and $\mathbf{P}$ which are nearly identical in the beginning and towards the end, the intermediate zone is characterized by increasing differences between the two.

A plot of the cumulative measure μ calculated as the total information generated up to a time epoch (Section 2.3) is shown in Fig. 3(ii). This also shows a gradual increase in accumulated information leading to a catastrophic change at $v = 0.320$ after which the measure μ is stable showing no further rise as the system shows no further change of characteristic behavior resulting in no further growth of anomalies. It is believed that to further investigate the behavior of the system after the transition at $v = 0.320$ a new partition of the phase might be required as the energy states might be too coarse for changes within the new phase to be picked-up, which could be the reason behind a low value of $\hat{I}$ after $v > 0.320$. The plot of μ is normalized to highlight relative change in magnitude. It must be noted that for this system the sudden transition from one mode of behavior to another has been treated as a catastrophic fault and changes leading to it as the growth of anomalies.

The results show that it is possible to detect incipient catastrophic change manifesting itself as a phase transition (discontinuity in $\hat{I}$) using the proposed methodology and only the time series of the data collected from the system. Also the gradual growth of the system towards that impeding failure can be tracked using the cumulative measure μ which shows a monotonic increase to a significant value much before the actual breakdown. Thus the methodology enables both diagnosis as well as prognosis by analyzing only the time-series data from a slowly changing dynamical system. This approach is now applied to a real mechanical set-up under cyclic load to monitor its degradation leading to total failure.

4. Experimental validation: fatigue damage detection

Mechanical engineering systems are usually subjected to loads of a cyclic nature, vibrations during operation for instance. This leads to cyclic stresses leading to fatigue failure which is one of the prominent causes of mechanical systems degradation.

The mechanical set-up used here for anomaly detection consists of a uniaxial fatigue testing machine designed to operate both under load and strain control. It can cycle at a maximum frequency of 12.5 Hz. The set-up consists of the following components.

(i) *Fatigue test unit*: Driven by a MTS hydraulic unit, the fatigue test unit is operated using computer-controlled electro-hydraulic servo-valves. The feedback system consists of a load-cell and an extensometer.

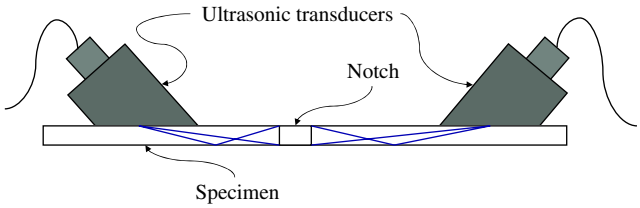

Fig. 4. Ultrasonics.

(ii) *Traveling optical microscope*: A traveling optical microscope is used to monitor visually the possible locations of crack initiation. Also, it provides direct measurements of the visible part of a crack when it develops. The resolution of the optical microscope is about 2 μm at a working distance of 10–35 cm.

(iii) *Ultrasonics*: The ultrasonic system essentially has two piezoelectric transducers which work in a pitch-catch mode. One of the transducers emits a high frequency (10 MHz sine wave) ultrasonic pulses which travel through the specimen and are collected at the other end of the notch by the receiver transducer. The transducers are placed so that a significant portion of the ultrasonic energy is sent through the region of expected crack propagation and received on the other side, see Fig. 4. Ultrasonic signal collected at the receiver end is analyzed by the method proposed in this paper, results from the experiment are presented in Section 4.2. A detailed description of the apparatus and its design specifications are provided in [33].

4.1. Experimental procedure

Fatigue tests have been conducted on side-notched 7075-T6 aluminum alloy specimen at a constant amplitude sinusoidal load for low-cycle fatigue, where the maximum and minimum loads were maintained at 87 and 4.85 MPa, respectively. The specimen is 3 mm thick and 50 mm wide with a side notch of 1.58 mm diameter and 4.57 mm length. The purpose of the notch is to increase the stress concentration and hence initiate a fatigue crack at notch end. For low-cycle fatigue the stress amplitude at the crack tip, including the stress concentration, is sufficiently high to observe the elasto-plastic behavior in the specimens under cyclic loading. As the specimen is loaded under cyclic stress, the behavioral characteristics of the specimen change with the accumulation of inelastic deformation near the notch end. Ultrasonic waveforms traveling through the specimen pick-up these changes and serve as an appropriate sensor time series for analysis using the proposed method. Each *epoch* consists of 50 load cycles in the *fast time scale* during which data are collected from the experiment. The time series of ultrasonic data is preprocessed using wavelet transforms using a Gaussian basis function [29]; maximum entropy partitioning described in Section 2.1 is used to partition the observed space into eight segments. The assumption, as mentioned in Section 2, is that the system remains statistically stationary during each *epoch* and anomalies evolve in the *slow time scale* τ, growing from one *epoch* to the other. The results and discussion of the conducted experiment are presented in the next subsection.

4.2. Results

Figs. 5 and 6 show the results of the analysis of ultrasonic data from the fatigue experiment. In Fig. 5 the images are from the microscope showing the condition of the material surface around the notch at four different points in the life-time of the specimen (cases a–d). The histograms in the corresponding bar charts represent the actual probability distribution $\hat{\mathbf{p}}$ (blue) and the canonical distribution $\mathbf{P}$ (yellow) of the eight cells or energy levels.

In the condition shown in case (a) the specimen is in the nominal condition, the corresponding image shows the initial condition of the notch site, this is taken at around 3000 cycles when any possible transients in the loading pattern would have died down. It can be seen that the notch site is healthy with no sign of any surface cracks. First appearance of a surface crack as seen by the microscope is at around 34,000 cycles shown in case (b), the approximate size of the crack at this point is $\approx 200\,\mu$m. Beyond this point the fatigue crack begins to grow rapidly, the state of the specimen at around 42,000 cycles is shown in case (c). Growth is continued eventually leading to total failure at $\approx 57,000$ cycles depicted in case (d). Histograms corresponding to the images from the microscope show the evolution of the probability vectors $\hat{\mathbf{p}}$ and $\mathbf{P}$, these in essence contain compressed information characteristic to the system at each point in time which is extracted using $\hat{I}$.

Fig. 6(i) shows the trend followed by $\hat{I}$ with load cycles. This plot can be divided into two regions—the first region from around 3000 cycles to 34,000 cycles (cases a and b) is the crack incubation pase, while the next region is of crack propagation starting from 34,000 cycles to end of life at 57,000 (from cases b to d). In the incubation phase $\hat{I}$ shows a small increase in value up till before the point where the crack is seen visually for the first time (case b). Whereas, the portion of the curve corresponding to propagation phase is marked by a peak formation in $\hat{I}$ attaining its highest value at $\approx 42,000$ cycles (case c). This is due to a rapid increase in information gained in each time epoch in the first half of propagation phase

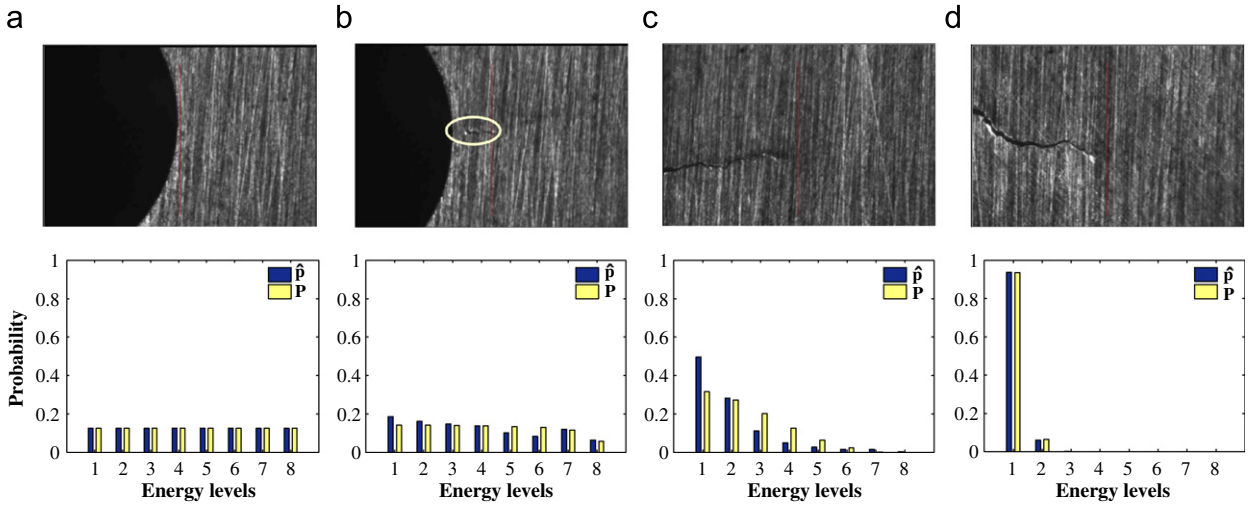

Fig. 5. Crack propagation: microscope images and probability histograms. (a) 3000, (b) ≈ 34,000, (c) ≈ 42,000, (d) ≈ 57,000.

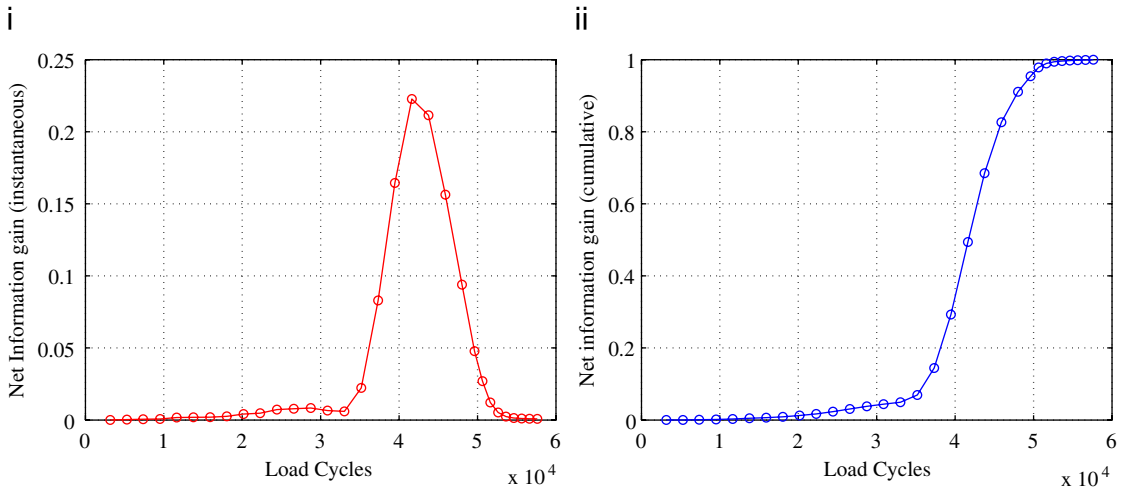

Fig. 6. Net information gain (i) instantaneous $\hat{I}(\hat{\mathbf{p}})$ and (ii) cumulative μ.

when the system is organizing towards failure, but with further growth of anomalies beyond this point (case c) the amount of new information about the impeding failure decreases with each time epoch as by now the event of failure is almost certain. This decreasing trend slowly leads towards the end of life shown in case d where the amount of new information contained in the data for the given macro-behavior is minimal, as at this point the system has completely degraded and is stable in that state. This can also be understood from the histograms shown in Fig. 5 where the initial and final probability distributions are almost similar while they differ the most during the growth phase showing maximum information gain in that region. This change from nominal behavior to the final degraded state can be seen as a phase change of a higher order in the thermodynamic sense.

A plot of measure μ quantifying cumulative damage is shown in Fig. 6(ii). This plot represents the accumulation of anomalies within the system up to a given load cycle. The gradual increase in μ from nominal state in case a to the state shown in case b is a precursor to a sudden change of behavior approaching in the near future. The cumulative measure rises quickly during the crack growth phase and becomes saturated towards the end when most of the damage has already occurred and there is no new degradation. When the anomaly is minimal, even a small increment in damage is a rather large degradation. However, towards the end of the specimen's fatigue life, further accumulation of anomalies leads to saturation of the cumulative measure. This trend is correctly captured by the cumulative measure plot of μ against load cycles. The plot is normalized to highlight the relative change of magnitude in μ. It is noteworthy that μ is essentially the area under the curve of $\hat{I}$ so while instantaneous degradation can be monitored using $\hat{I}$, μ tracks the extent of total damage.

5. Summary, conclusions, and future research

Thermodynamic formalism has been used to construct a dynamic data-driven method for anomaly detection in complex systems. The anomaly detection method is built upon the fundamental principles of information theory and statistical mechanics. A new measure of information gain is constructed and its properties are identified. The efficacy of the measure is first tested on a known nonlinear system exhibiting bifurcation and chaos. It is shown that this measure can successfully monitor slowly evolving growth of anomaly as well as those leading to abrupt changes. The validated anomaly detection procedure is then applied to fatigue damage monitoring under cyclic load, where the underlying structure of the damage growth process is unknown. The results show that it is possible to detect the appearance of small cracks (200 μm) and monitor the growth of fatigue damage. The growth phase of the crack was also successfully monitored by the cumulative measure μ which reached a saturated value corresponding to the terminal state of total failure. The measures of information gain formulated in this paper were shown to be useful to this effect for two systems considered here. Their application in the other fields of study can be of significant use.

It is concluded that thermodynamic formalism would facilitate formation of links between the micro- and macro-behavior of multi-time-scale dynamical systems and provide a computationally tractable measure of deviations from the nominal behavior. This information is useful for prediction and mitigation of forthcoming failures. The reported work is a step towards building a reliable instrumentation system for early detection and mitigation of faults in human-engineered complex systems, such as electric power generation, petrochemical, and networked transportation. Further research is necessary before their applications can be realized in industry.

5.1. Recommendations for future research

Although thermodynamic formalism, as a potential tool of analyzing dynamical systems, has been studied over the last few decades, further theoretical and experimental research is needed to formulate concepts and algorithms for application to engineering problems. Pertinent research directions in this area are listed below.

- Incorporation of more than one macro-variable in the thermodynamic formalism to represent dynamical systems (i.e., the macro-parameters are vectors as opposed to being scalars).
- Appropriate selection of the function φ in Eq. (1) to define the extensive parameters (e.g., energy, and volume) corresponding to macroscopic thermodynamic states.
- Re-partitioning of the phase space upon detection of a phase transition.
- Investigation of new partitioning methods such as those based on Hilbert transform [34].
- Further experimentation on both the current and new laboratory apparatuses for repeatability analysis and solution of the inverse problem [35,36].

Appendix A. Thermodynamic equilibrium ensembles

This appendix elaborates basic statistical mechanical concepts presented in Section 2.2. If the observed macro-variable is not a vector and more than one macroscopic averages are needed to characterize the dynamical system, then the statistics of micro-motion can be estimated as follows [21].

Let M^σ be one such macroscopic variable and $\mathbf{M}^\sigma$ be its mean value observed over an *epoch*

$$\mathbf{M}^\sigma = \sum_{i=1}^{m} p_i M_i^\sigma \tag{A.1}$$

where $\mathbf{p}$ is the probability vector satisfying

$$1 = \sum_{i=1}^{m} p_i \quad \text{and} \quad p_i \geqslant 0 \ \ \forall i \tag{A.2}$$

For an unbiased guess of $\mathbf{p}$, the Shannon entropy is maximized (Eq. (5)) such that constraints given by Eqs. (A.1) and (A.2) are satisfied. Therefore, for an infinitesimal variation δp_i and using Ψ and β_σ as Lagrange multipliers, it follows that

$$\sum_{i=1}^{m} (\ln p_i - \Psi + \beta_\sigma M_i^\sigma)\delta p_i = 0 \tag{A.3}$$

Since all δp_i's are arbitrary the summand vanishes and we have desired distribution $\mathbf{P} = [P_1, \ldots, P_m]$ called the *generalized canonical distribution* or the *Gibbs distribution* given as

$$P_i = \exp(\Psi - \beta_\sigma M_i^\sigma) \tag{A.4}$$

If there are more than one macro-parameter M^σ, the term $\beta_\sigma M_i^\sigma$ becomes a sum over σ, often written using the Einstein's notation for summation. As opposed to the product βE_i appearing in Eq. (8), we now have the sum $\beta_\sigma M_i^\sigma$ appearing the expression for P_i. For special choices of M^σ the distribution $\mathbf{P} = [P_1, \ldots, P_m]$ takes on the following well known forms.

For $M^\sigma = E$, the distribution $\mathbf{P}$ is called the *canonical distribution* as used in Eq. (8)

$$P_i = \exp[\beta(F - E_i)] \tag{A.5}$$

Here $\beta = (k_b T)^{-1}$ is the inverse of temperature and k_b is the Boltzmann constant. $F = (1/\beta)\Psi(\beta)$ is the *Helmholtz free energy*.

When the macro-variables are energy and volume we have

$$P_i = \exp[\beta(G - E_i - \Pi V_i)] \tag{A.6}$$

called the *pressure ensemble* and $G = (1/\beta)\Psi(\beta)$ is the *Gibbs free energy*. For a given energy, volume and number of particles we have

$$P_i = \exp[\beta(\Omega - E_i - \Pi V_i + \mu N_i)] \tag{A.7}$$

called the *grand canonical ensemble* and μ is the *chemical potential*.

Using the generalized canonical distribution the Shannon entropy can now be written as

$$\begin{aligned} S = & -\sum_{i=1}^{m} P_i \ln P_i \\ = & -\Psi + \beta_\sigma \mathbf{M}^\sigma \end{aligned} \tag{A.8}$$

where $\mathbf{M}^\sigma$ is the observed mean value of the macro-variable M^σ. Also, the quantity $\sum_{i=1}^{m} \exp(-\beta_\sigma M_i^\sigma)$ is called the *partition function* Z and using $\sum P_i = 1$ we can relate Ψ and Z as

$$\Psi = -\ln Z \tag{A.9}$$

Appendix B. Phase-space partitioning of dynamical systems

This appendix summarizes the basic concepts of phase-space partitioning of a dynamical system based on wavelet transformation of the time-series data.

B.1. Wavelet analysis

The choice of appropriate parameters for computing the wavelet transform of a signal is necessary for data conditioning. Parameters of interest are the *shape* and the *scales* of the wavelet basis function.

Since the wavelet coefficients are computed using time translations of a wavelet basis function at a given scale. The wavelet basis function is chosen to closely match the shape of an appropriately scaled and translated signal. Adaptive wavelet design can also be used and is an area of current research [37] where the basis function is engineered to specifications.

Given a wavelet basis function, the frequency with the maximum modulus in its Fourier transform is called the *center frequency* F_c. For a given scale α the pseudo-frequency f_p^α of the wavelet at that scale is defined as

$$f_p^\alpha = \frac{F_c}{\alpha \Delta t} \tag{B.1}$$

where Δt is the sampling interval. The power spectral density of the signal contains information about the frequency contents of the signal, coupled with Eq. (B.1) this information can be used to select the relevant scales for the wavelet analysis. When the locally dominant frequencies in the signal correspond to the pseudo frequency f_p^α of the wavelet at a scale α, it is reflected as larger values of the wavelet coefficients. Thus, the transient features of the signal can be localized in time and frequency, and extracted using wavelet analysis. A detailed discussion of the wavelet-based partitioning can be found in [11].

B.2. Partition size

The choice of adequate number of cells m for partitioning is determined using the *entropy rate* $h(\ell)$, where the entropy $H(\ell)$ of the system is computed as a function of the number of states ℓ, given a sufficiently large data set. The entropy rate is defined as

$$h(\ell) = H(\ell) - H(\ell - 1) \tag{B.2}$$

Note: the entropy rate is computed as a function of ℓ and not time.

For an ideal (i.e., noise free) signal, the entropy rate would eventually vanish as the number of states is increased to model the dynamical system. However, in the presence of noise, the entropy rate would always be nonzero (e.g., for white

noise, $h(\ell) = \log(\ell/(\ell-1)) \geqslant 0$ for any finite ℓ). A threshold ε_h is used to decide the number of states m such that [11]

$$h(\ell) = H(\ell) - H(\ell-1) \leqslant \varepsilon_h \quad \forall \ell \geqslant m \tag{B.3}$$

B.3. Data window size

When working with finite data sequences, a *stopping rule* [31] is necessary to find the lower bound on the number of data points. Given a sampling rate, the number of data points is equivalent the time window length (i.e., *epochs*) and has a direct effect on the stationarity of the probability vector $\hat{\mathbf{p}}$. Given the partition size m and a threshold η, a lower bound r_{stop} of the required number r of data points is derived as follows.

In the computation of the probability vector $\mathbf{p}$, the convergence criterion is satisfied when the absolute error between successive iterations is given as

$$\|(\mathbf{p}(r) - \mathbf{p}(r+1))\|_\infty \leqslant \frac{1}{r} \tag{B.4}$$

where $\| \bullet \|_\infty$ is the max norm of the finite-dimensional vector $\bullet$. To calculate the stopping point r_{stop}, a tolerance of η, where $0 < \eta \ll 1$, is specified for the relative error such that

$$\frac{\|(\mathbf{p}(r) - \mathbf{p}(r+1))\|_\infty}{\|(\mathbf{p}(r))\|_\infty} \leqslant \eta \quad \forall r \geqslant r_{\text{stop}} \tag{B.5}$$

The objective is to obtain the least conservative estimate for r_{stop} such that the dominant elements of the probability vector have smaller relative errors than the remaining elements. Since the minimum possible value of $\|(\mathbf{p}(r))\|_\infty$ for all r is $1/m$, where m is the dimension of $\mathbf{p}(r)$, the least of most conservative values of the stopping point is obtained from Eqs. (B.4) and (B.5) as

$$r_{\text{stop}} \equiv int\left(\frac{m}{\eta}\right) \tag{B.6}$$

where $int(\bullet)$ is the integer part of the real number $\bullet$.

References

[1] A. Marcos, S. Ganguli, G. Balas, Application of h-infinity fault detection and isolation to a Boeing 747-100/200 aircraft, in: Guidance, Navigation, and Control Conference and Exhibit, AIAA, Monterey, CA, USA, 2002.
[2] A. Widodo, B.-S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mechanical Systems and Signal Processing 21 (6) (2007) 2560–2574.
[3] M. Misra, H.H. Yue, S.J. Qin, C. Ling, Multivariate process monitoring and fault diagnosis by multi-scale PCA, Computers and Chemical Engineering 26 (9) (2002) 1281–1293.
[4] D. Chelize, J.P. Cusumano, A dynamical systems approach to failure prognosis, Journal of Vibration and Acoustics 126 (1) (2004) 2–8.
[5] J.M. Nichols, M.D. Todd, M. Seaver, L.N. Virgin, Use of chaotic excitation and attractor property analysis in structural health monitoring, Physical Review E 67 (1) (2003) 016209.
[6] B.I. Epureanu, S.-H. Yin, M.M. Derriso, High-sensitivity damage detection based on enhanced nonlinear dynamics, Smart Materials and Structures 14 (2) (2005) 321–327.
[7] W.J. Wang, Z.T. Wu, J. Chen, Fault identification in rotating machinery using the correlation dimension and bispectra, Nonlinear Dynamics 25 (4) (2001) 383–393.
[8] M. Markou, S. Singh, Novelty detection: a review—part 2: neural network based approaches, Signal Processing 83 (12) (2003) 2499–2521.
[9] M. Markou, S. Singh, Novelty detection: a review—part 1: statistical approaches, Signal Processing 83 (12) (2003) 2481–2497.
[10] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Processing 84 (7) (2004) 1115–1130.
[11] V. Rajagopalan, A. Ray, Symbolic time series analysis via wavelet-based partitioning, Signal Processing 86 (11) (2006) 3309–3320.
[12] C. Daw, C. Finney, E. Tracy, A review of symbolic analysis of experimental data, Review of Scientific Instruments 74 (2) (2003) 915–930.
[13] S. Chin, A. Ray, V. Rajagopalan, Symbolic time series analysis for anomaly detection: a comparative evaluation, Signal Processing 85 (9) (2005) 1859–1868.
[14] S. Gupta, A. Ray, E. Keller, Symbolic time series analysis of ultrasonic data for early detection of fatigue damage, Mechanical Systems and Signal Processing 21 (2) (2007) 866–884.
[15] R. Pathria, Statistical Mechanics, second ed., Butterworth-Heinemann, Oxford, England, 1996.
[16] T. Cover, J. Thomas, Elements of Information Theory, Wiley, NY, USA, 1991.
[17] S. Gupta, A. Srivastav, A. Ray, Pattern identification in complex systems: a statistical thermodynamic approach, in: International Conference on Manufacturing Science and Engineering, ASME, Ypsilanti, MI, USA, 2006.
[18] A. Srivastav, A. Ray, S. Gupta, Irreversibility-based measure of slowly evolving anomalies, in: American Control Conference, New York City, NY, 2007.
[19] J. Thompson, H. Stewart, Nonlinear Dynamics and Chaos, Wiley, Chichester, UK, 1986.
[20] V.L. Berdichevsky, Thermodynamics of Chaos and Order, Addison Wesley Longman, Essex, England, 1997.
[21] C. Beck, F. Schlögl, Thermodynamics of Chaotic Systems: An Introduction, Cambridge University Press, Cambridge, UK, 1993.
[22] D. Ruelle, Thermodynamic Formalism, second ed., Cambridge University Press, Cambridge, UK, 2004.
[23] M. Paluš, V. Komárek, Z.c.v. Hrnčíř, K. Štěrbová, Synchronization as adjustment of information rates: detection from bivariate time series, Physical Review E 63 (4) (2001) 046211.
[24] J.M. Nichols, M. Seaver, S.T. Trickey, M.D. Todd, C. Olson, L. Overbey, Detecting nonlinearity in structural systems using the transfer entropy, Physical Review E (Statistical, Nonlinear, and Soft Matter Physics) 72 (4) (2005) 046217.
[25] T. Schreiber, Measuring information transfer, Physical Review Letters 85 (2) (2000) 461–464.
[26] M. Buhl, M. Kennel, Statistically relaxing to generating partitions for observed time-series data, Physical Review E 71 (4) (2005) 046213.
[27] D. Chelize, J.P. Cusumano, A. Chatterjee, A dynamical systems approach to damage evolution tracking, part 1: description and experimental application, Journal of Vibration and Acoustics 124 (2) (2002) 250–257.

[28] R. Badii, A. Politi, Complexity: Hierarchical Structures and Scaling in Physics, Cambridge University Press, Cambridge, UK, 1997.
[29] S. Mallat, A Wavelet Tour of Signal Processing, second ed., Academic Press, CA, USA, 1998.
[30] Z.K. Peng, F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mechanical Systems and Signal Processing 18 (2) (2004) 199–221.
[31] S. Gupta, A. Ray, Symbolic dynamic filtering for data-driven pattern recognition, in: E.A. Zoeller (Ed.), Pattern Recognition: Theory and Application, Nova, Science Publisher, Hauppage, NY, USA, 2007 (Chapter 2).
[32] J.H. Kim, J. Stringer (Eds.), Applied Chaos, Wiley, NY, USA, 1992.
[33] E. Keller, A. Ray, Real-time nondestructive evaluation of mechanical structures, Structural Health Monitoring 2 (3) (2003) 191–203.
[34] A. Subbu, A. Ray, Space partitioning via Hilbert transform for symbolic time series analysis, Applied Physics Letters 92 (8) (2008) 084107.
[35] S. Gupta, A. Ray, Real-time fatigue life estimation in mechanical structures, Measurement Science and Technology 18 (7) (2007) 1947–1957.
[36] V. Rajagopalan, S. Chakraborty, A. Ray, Estimation of slowly-varying parameters in nonlinear systems via symbolic dynamic filtering, Signal Processing 89 (2) (2008) 339–348.
[37] W. Sweldens, The lifting scheme: a construction of second generation wavelets, SIAM Journal of Mathematical Analysis 29 (2) (1998) 511–546.