

Estimation of Multiple Parameters in Dynamical Systems

Chinmay Rao Kushal Mukherjee Soumik Sarkar Asok Ray[†]
crr164@psu.edu kum162@psu.edu szs200@psu.edu axr2@psu.edu

The Pennsylvania State University
University Park, PA 16802

Abstract—This paper addresses the problem of multiple parameter estimation in dynamical systems, where the solution algorithm is built upon the principles of extracting statistical information contents or patterns in the framework of Symbolic Domain Filtering. The proposed algorithm has been tested for estimation of two slowly varying parameters in an active electronic system that is constructed in the classical Duffing equation setting.

Index Terms—Symbolic Dynamics, Parameter Estimation, Tsallis Thermodynamics

I. INTRODUCTION

RECENT research has explored the problem of anomaly detection in dynamical systems based on Symbolic Dynamic Filtering (*SDF*) [1]–[3]. Component level fault isolation is of practical significance for health monitoring of dynamical systems that are composed of mutually interacting subsystems. This is especially important if tractable models are not available for individual components that are richly interconnected, physically, as well as through the use of feedback control loops. The rationale is that degradation in any one component may affect the input condition for other components. Thus, detection and isolation of anomalous behavior in multiple components and estimation of the evolving fault magnitude for the purpose of prognoses and health monitoring pose a challenging system identification problem. There are several non-linear system identification techniques available, such as Bayesian filtering and neural networks [4].

The key problem, addressed in this paper, is identification of statistical patterns, which represent evolving dynamical behavior of the system. The pattern identification problem is formulated in terms of observation-based estimation of the process variables. In other words, in the absence of a feasible mathematical model, the inherent dynamics of a nonlinear dynamical system are inferred from time series data generated from sensing devices that are sensitive enough to capture the essential information on the system dynamics. Information-based inference of the underlying process becomes a formidable task. These complexity issues in the information-based inference of the underlying process have motivated the study of dynamical systems from the perspectives of Statistical Mechanics [5], [6].

This work has been supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant No. W911NF-07-1-0376 and by NASA under Grant No. NNX07AK49A.

II. SINGLE PARAMETER ESTIMATION PROBLEM

This section explains how anomaly detection algorithms are constructed for estimating a single parameter in a complex dynamical system under the symbolic dynamic filtering (*SDF*) framework. A brief description of the construction of the anomaly measure and the parameter estimation from this anomaly measure for the Duffing equation follows [7].

A. Construction of Anomaly Measure using Symbolic Dynamics

The following steps, summarize the procedure of *SDF* for anomaly detection.

- **Time series data acquisition** on the fast scale from sensors and/or analytical measurements. Data sets are collected at different slow time epochs $t_0, t_1, t_2, \dots, t_k, \dots$
- **Generation of wavelet transform coefficients** [8], obtained with an appropriate choice of the wavelet basis and scales [2].
- **Partitioning** [2] of the wavelet space at the nominal condition at time epoch t_0 . Each segment of the partitioning is assigned a symbol from the alphabet Σ .
- **Construction of a finite state automaton** at time epoch t_0 (nominal condition) from alphabet size $|\Sigma|$ and window length D . The structure of the finite state machine is fixed for subsequent slow time epochs $\{t_1, t_2, \dots, t_k, \dots\}$.
- **Calculation of the state probability vectors** $\mathbf{p}^0, \mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^k, \dots$. The probability distribution \mathbf{p}^0 of patterns is recursively computed as an approximation of the natural invariant density [5] of the dynamical system at the slow time epoch t_0 . Subsequently $\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^k, \dots$ at slow time epochs, $t_1, t_2, \dots, t_k, \dots$ are computed from the respective symbolic sequences using the finite state machine constructed at time epoch t_0 .
- **Computation of scalar anomaly measures** $\mu^1, \mu^2, \dots, \mu^k, \dots$ based on evolution of these probability vectors and by defining an appropriate scalar distance function $\mu_k = d(\mathbf{p}^k, \mathbf{p}^0)$ with respect to the nominal condition [1].

B. Single Parameter experiment on an electronic system

An electronic system apparatus is constructed, as described in [7], based on the Duffing equation [9] that is a

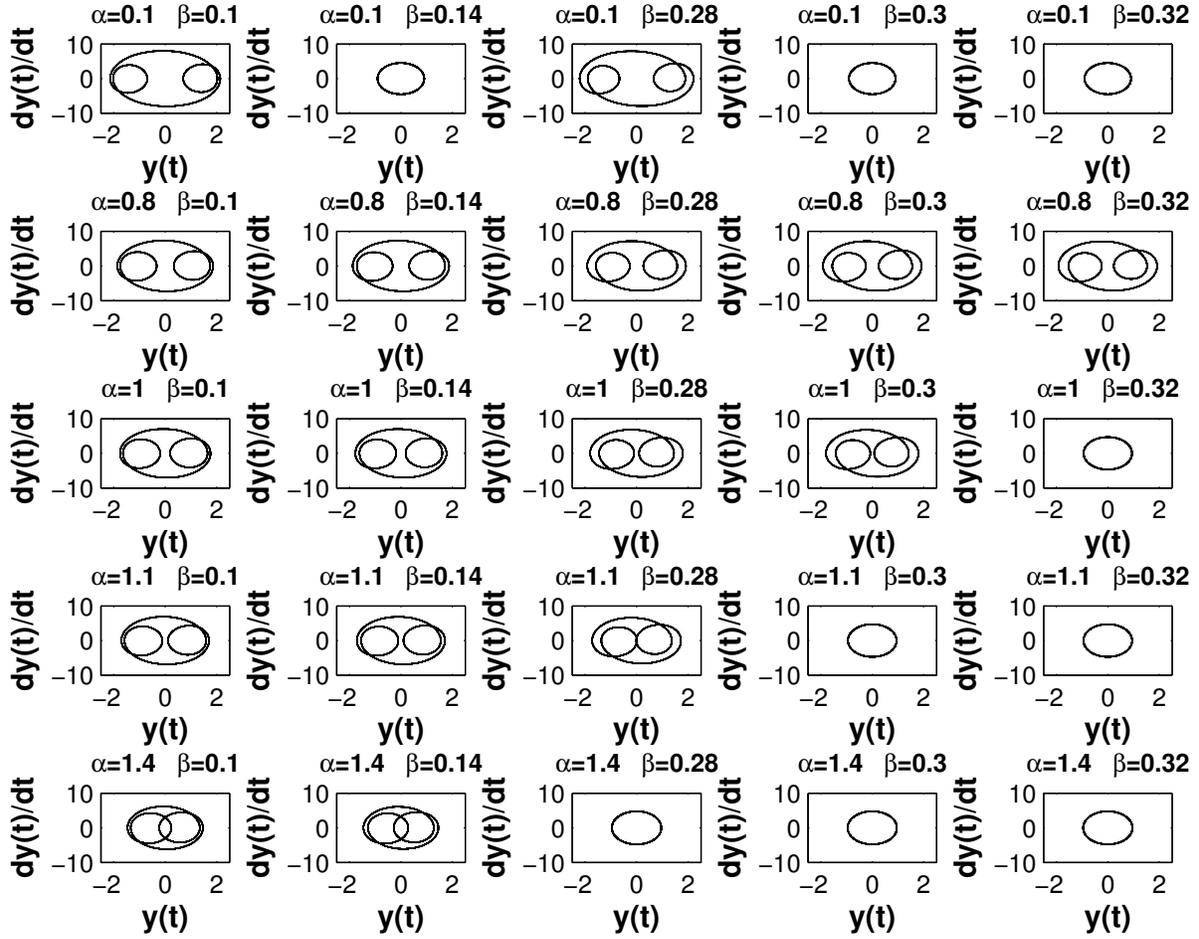


Fig. 1. Phase Plots for the Multi-Parameter Duffing Experiment

second-order forced differential equation with a cubic non-linearity:

$$\frac{d^2x(t)}{dt^2} + \beta \frac{dx}{dt} + x(t) + x^3(t) = A \cos(\omega t) \quad (1)$$

The dissipation parameter $\beta(t_s)$, realized as a resistance in the circuit, varies with the slow time t_s and is treated as a constant in the fast time scale at which the dynamical system is excited. The goal is to detect, at an early stage, changes in $\beta(t_s)$, which are associated with the anomaly. For illustration purposes, we show the response of a stimulus with amplitude $A = 22$ and frequency $\omega = 5$.

C. Inverse Problem: Single Parameter Solution

This section focuses on the inverse problem of parameter estimation based on computed values of the deviation measure. The parameter is a slowly varying random process and is therefore assumed to be a random variable at each slow time epoch, for which the deviation measures are the only observables. To account for the inherent uncertainties in the system components and to ensure robust estimation, a large number of experiments are performed and the deviation measures are calculated from observed time series data during every experiment, with the objective of estimating the unknown parameter.

The range of the computed deviation measure is discretized into finitely many levels. A pattern matrix is created where each column represents the spread of the parameter for a particular value of the deviation measure. A statistical distribution is hypothesized for the spread of the parameter and the goodness-of-fit of the hypothesized distribution is assessed with χ^2 and Kolmogorov-Smirnov tests.

III. THE MULTI-PARAMETER ESTIMATION PROBLEM

A. Multiple Parameter Experiment

The Duffing equation is a second-order forced differential equation with a cubic non-linearity. It is given by

$$\frac{d^2x(t)}{dt^2} + \beta \frac{dx}{dt} + \alpha_1 x(t) + x^3(t) = A \cos(\omega t) \quad (2)$$

In this experiment, the dissipation parameters are chosen as β and α_1 . The input stimulus are chosen as $A = 5$ and $\omega = 5$. The stationary behavior of the system is obtained with several combinations of values, with β ranging from 0.10 to 0.40, and α_1 ranging from 0.10 to 1.50. The phase plots are shown in Figure 1. The third row of plots corresponds to a value of $\alpha_1 = 1.0$ which is exactly the same as considered in Section II-B. It can be seen that increasing (decreasing) the value of α_1 causes an early (late) onset of bifurcation.

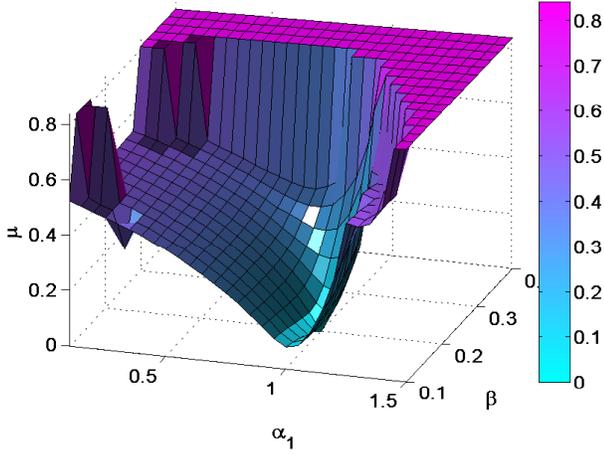


Fig. 2. Three Dimensional view of Anomaly Measure μ

Also, a bifurcation is associated simply with a rise in α_1 for low values of β as can be seen in the first row of plots. It is also clear that there is very little visible difference in most plots before the onset of bifurcation. The challenge of this experiment is to determine the values of *both* α_1 and β by looking at data gathered at a slow time scale.

B. Formulation of the Problem Statement

The method outlined in Section II-C cannot directly be extended to a multiple parameter framework. That is, the anomaly measure μ obtained from a single time series cannot always be used to isolate and estimate multiple faults in a system. Consider the Duffing equation elaborated in Section III-A. The anomaly measure μ is obtained for each pair of (β, α_1) values, for one single run of the system and a three dimensional plot is obtained as shown in Figure 2, with α_1 on the x-axis, β on the y-axis and the anomaly measure μ on the z-axis. Figure 3 shows a contour plot with each line joining points with the same value of anomaly measure μ . The value of μ is indicated by the color of the line, with a colorbar to the right of the plot. The nominal condition is represented by $\alpha_1 = 1, \beta = 0.1$ which corresponds to an anomaly measure of 0. It can be seen that μ increases as the system deviates in either direction from the nominal condition, rising sharply at bifurcation. Unlike the single parameter case, the anomaly curves are not normalized.

It is clear from Figure 3 that a single value of μ would contribute to infinitely many combinations of (β, α_1) . For example, taking a section along $\mu = 0.4$ gives a contour. Now, every point (β, α_1) along this contour yields the same value of μ and hence the current information does not help in isolating the exact values of (β, α_1) that describe the system at the present case.

IV. THERMODYNAMIC FORMALISM FOR PARAMETER ESTIMATION

In order to solve this problem, the principles of thermodynamics and statistical mechanics [5] are invoked. In

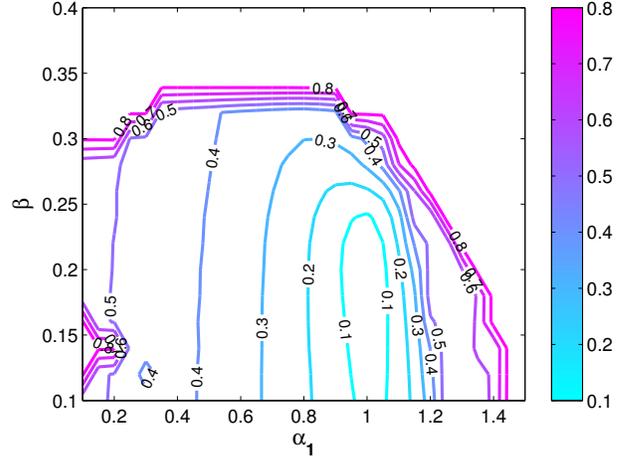


Fig. 3. Contour Plot of Anomaly Measure μ

statistical mechanics, a few macroscopic parameters (e.g. pressure and temperature) are used to describe the intrinsic dynamics of the entire system in terms of the estimates derived from the distribution of the elementary particles in various micro states. In the same fashion, the behavior of a dynamical system can be investigated both from microscopic and macroscopic points of view. In the study of a dynamical system, the measured time series data of the observable parameters on the fast time scale can be analyzed to generate the pattern vectors in terms of the probability distributions, which can be used to describe the macroscopic or global behavior of the system at a particular slow time epoch. The information derived from these pattern vectors can be further compressed into a few macroscopic parameters such as entropy, Kullback distance, and Euclidean norm.

Statistical Mechanics \Rightarrow distribution of microstates \rightarrow macroscopic properties

Dynamical System \Rightarrow pattern vectors from time series data \rightarrow System Parameters

A. Escort Probabilities and Distributions

Escort distributions [5] scan the attributes of the original distribution, while describing the features of a non-linear dynamical system. [10]. Let $\{p_i\}$ be the original distribution. Then its escort is given by:

$$P_i = \frac{\phi(p_i)}{\sum_j \phi(p_j)} \quad (3)$$

where ϕ is a positive function.

This equation comes about when we consider that the Renyi information is a monotonically increasing function of β [5]. An important case occurs if $\phi(s) = s^q$, for $0 < s \leq 1$ and $q > 0$ then $P_i^{(q)} \equiv P_i$ and

$$P_i^{(q)} = \frac{(p_i)^q}{\sum_j (p_j)^q} \quad (4)$$

Expectations with respect to the original distribution p are denoted as \mathbb{E}_p . Also, expectations with respect to the escort distribution $P^{(q)}$ are denoted as \mathbb{F}_q . More formally,

$$\mathbb{E}_p f = \int_{\Omega} d\mu(x) p(x) f(x) \quad (5)$$

and

$$\mathbb{F}_q f = \int_{\Omega} d\mu(x) P^{(q)}(x) f(x) \quad (6)$$

Now, $p_i \xrightarrow{q} P_i^{(q)}$ can be regarded as a transformation.

$$P_i^{(q)} \xrightarrow{r} = \frac{(p_i)^{qr}}{\sum_j (p_j)^{qr}} = P_i^{(qr)} \quad (7)$$

This transformation forms a one-parameter Abelian group with the identity transformation corresponding to the order unity. The parameter is obviously, q .

In the context of Symbolic Dynamics, it is proposed that the State Probability Vector corresponds to the original distribution, $\{p_i\}$, while the escort distribution is $P_i^{(q)} = \frac{(p_i)^q}{\sum_j (p_j)^q}$ where q is defined as the Tsallis degree of non-extensivity of the complex dynamical system. In the two cases considered in this paper, it is assumed that $q = 1$ and the original distribution and the escort distribution are the same.

B. Parameter Estimation using Escort Probabilities

The concept of Fisher information is introduced in a thermodynamic sense, which is then applied to the problem of parameter estimation. The discussion follows the principles embodied in [5], [10]. In statistics and information theory, the Fisher information (denoted $\mathcal{I}(\beta)$) is the variance of the score. The Fisher information is the amount of information that an observable random variable X carries about an unknown parameter β upon which the likelihood function of X , $\mathcal{I}(\beta) = f(X; \beta)$, depends. The likelihood function is the joint probability of the data, the X s, conditional on the value of β , as a function of β . Since the expectation of the score is zero, the variance is simply the second moment of the score, the derivative of the log of the likelihood function with respect to β .

$$\mathcal{I}(\beta) = \mathbb{E} \left\{ \left[\frac{\partial}{\partial \beta} \ln f(X; \beta) \right]^2 \middle| \beta \right\}, \quad (8)$$

This implies $0 \leq \mathcal{I}(\beta) < \infty$.

To introduce a thermodynamic formalism based definition of the Fisher information, it is necessary to first define Kullback information and the Kullback Distance.

The Kullback Liebler relative entropy is defined for two distributions π and π^a as:

$$K[\pi \parallel \pi^a] = \sum_i \pi_i \ln \frac{\pi_i}{\pi_i^a} \quad (9)$$

This is positive definite and vanishes only if $\pi_i = \pi_i^a \forall i$.

The Kullback Liebler divergence is defined for the same two distributions as:

$$D[\pi, \pi^a] = K[\pi \parallel \pi^a] + K[\pi^a \parallel \pi] \quad (10)$$

Let π_i depend on a set of parameters \mathbf{q} . That is, let $\pi_i = \pi_i(\mathbf{q})$, where $\mathbf{q} = (q^1, q^2, \dots, q^n)$.

Let π_i^a represent $\pi_i(\mathbf{q} + d\mathbf{q})$. $D[\pi, \pi^a]$ is calculated as:

$$D[\pi, \pi^a] = \sum_{\mu, \nu=1}^n g_{\mu\nu}(\mathbf{q}) dq_{\mu} dq_{\nu} \quad (11)$$

where $g_{\mu\nu}$ is given as:

$$g_{\mu\nu}(\mathbf{q}) = \sum_i \frac{\partial_{\mu} \pi_i(\mathbf{q}) \partial_{\nu} \pi_i(\mathbf{q})}{\pi_i(\mathbf{q})} \quad (12)$$

$g_{\mu\nu}$ is defined as the Fisher Information, or the Fisher metric, and $\partial_{\mu} = \frac{\partial}{\partial q_{\mu}}$. \mathbf{q} supplies a local coordinate in the n -dimensional submanifold of the functional space of distributions. The Fisher metric is merely an induced metric on this manifold [5], [10].

V. MULTIPLE PARAMETER ESTIMATION FRAMEWORK

The principles detailed in section IV are now applied to the problem of multiple parameter estimation. The following analogies are proposed:

$$\pi \equiv \mathbf{p}^0 \quad (13a)$$

$$\pi^a \equiv \mathbf{p}^k \quad (13b)$$

$$\mathbf{q} \equiv \underline{\beta} \quad (13c)$$

It is proposed that the escort probabilities in Gibbs' extensive thermodynamics correspond to the state probabilities in symbolic dynamics. Just as the escort probabilities evolve under a thermodynamic process, the state probabilities evolve as the complex dynamical system is run. Also, the escort probabilities represent a map to the physical properties of the system, in a manner analogous to the state probabilities in *SDF*. The parameter set \mathbf{q} is considered analogous to $\underline{\beta}$ which can itself be considered to be a vector of system parameters i.e. $\underline{\beta} = [\beta_1, \beta_2, \dots, \beta_n]$

The two equations (11) and (12) can be combined to obtain:

$$D[\pi, \pi^a] = \sum_{\mu, \nu=1}^n \sum_i \frac{\frac{\partial}{\partial q_{\mu}} \pi_i(\mathbf{q}) \frac{\partial}{\partial q_{\nu}} \pi_i(\mathbf{q})}{\pi_i(\mathbf{q})} dq_{\mu} dq_{\nu} \quad (14)$$

The inner summation goes over the length of the $\pi_i(\mathbf{q})$ vector. If we invoke the analogy listed in (13), we can substitute $\pi^a \equiv \mathbf{p}^k$, and the summation will be over the number of symbols (alphabet size) i.e. $|\Sigma|$ used in symbolic dynamic filtering.

$$D[\mathbf{p}^k(q), \mathbf{p}^0(q)] = \sum_{\mu, \nu=1}^N \sum_i^{|\Sigma|} \frac{\frac{\partial}{\partial q_{\mu}} \mathbf{p}^k(q) \frac{\partial}{\partial q_{\nu}} \mathbf{p}^k(q)}{\mathbf{p}^k(q)} dq_{\mu} dq_{\nu} \quad (15)$$

The $\mathbf{p}^k(q)$ in the denominator serves as a normalizing constant. Now, q is replaced by $\underline{\beta}$. In order to do this, we

need to define how partial derivatives are computed in this setting. Define:

$$\frac{\partial \mathbf{p}_{\beta_1}^k}{\partial \beta_1} d\beta_1 = \mathbf{p}(\beta_1^{k+1}, \beta_2, \dots, \beta_n) - \mathbf{p}(\beta_1^k, \beta_2, \dots, \beta_n) \quad (16)$$

where $\mathbf{p}(\beta_1^k, \beta_2, \dots, \beta_n)$ represents the probability vector for the k th slow time epoch for β_1 and a nominal condition for $\beta_2 \dots \beta_n$.

Now, the distances are evaluated for all values of μ and ν over all state symbols. It is proposed that the probability of occurrence of a particular combination of $\beta_1, \beta_2, \dots, \beta_n$ are inversely proportional to the distance $D[\mathbf{p}^k(\beta), \mathbf{p}^0(\beta)]$.

When a time series data is obtained at the slow time scale, which depends on a series of β values, then it is initially passed through the symbolic dynamic filter. The probability sequence \mathbf{p}^k is obtained and the Kullback distance is determined for a set of points in a stored library. Depending on the location of this \mathbf{p}^k in a $|\Sigma|$ dimensional space, and its proximity to various points, the parameters $\underline{\beta}$ can be determined.

This approach can be viewed as determining contours for each element of \mathbf{p}^k . Let $\mathbf{p}^k = [\mathbf{p}_1^k, \mathbf{p}_2^k, \dots, \mathbf{p}_{|\Sigma|}^k]$ where the subscript denotes the energy states, and varies from 1 to $|\Sigma|$. The point of intersection of these $|\Sigma|$ number of contours then gives the point corresponding to $\underline{\beta}$ which is to be determined. A variation of this approach involves finding a probabilistic map for each symbol for all possible values of $\underline{\beta}$. Then, an overall probabilistic map is found by giving equal weightage to all symbols, and a most likely estimate of $\underline{\beta}$ can be determined.

The methods described in Section V are implemented on the Duffing Data with multiple parameters. Snapshots of the data have been presented in Figure 1. During training, test data is generated with values of β ranging from 0.10 to 0.40 in steps of 0.02, and values of α_1 from 0.01 to 1.50 in steps of 0.02. This data is then passed through an optimally constructed symbolic dynamic filter, and values of the probability vector $\mathbf{p}(\alpha_1, \beta)$ are obtained. These values are rounded off to the same level of precision considered while constructing the stopping rule for *SDF*. During testing, time series data are generated with process noise variance $w = 0.001$ and sensor noise variance $v = 0.01$, and unknown values of (α_1^a, β^a) . This data is then passed through the same symbolic dynamic filter, that was used during the forward problem and values of the probability vector $\mathbf{p}(\alpha_1^a, \beta^a)$ are obtained. The distances are computed, and for each $[\mathbf{p}_1^k, \mathbf{p}_2^k, \dots, \mathbf{p}_{|\Sigma|}^k]$, a contour (i.e. a set of candidate (α_1^a, β^a)) is obtained. An estimate of (α_1^a, β^a) is calculated from the intersection of all these contours. Two randomly chosen values of (α_1^a, β^a) are selected, one which is close to a value in the training set, and one that is relatively further away from values in the training set, and the results are shown below. Two implementations are shown, one which relies on direct contour intersection and one that relies on the probabilistic approach.

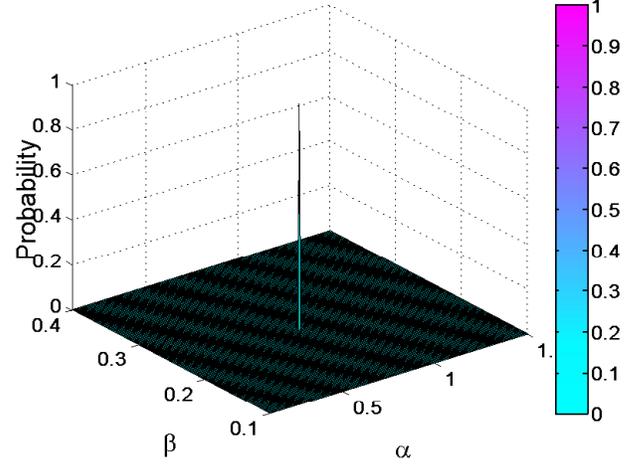


Fig. 4. Probability for various points for the two parameter Duffing estimation problem

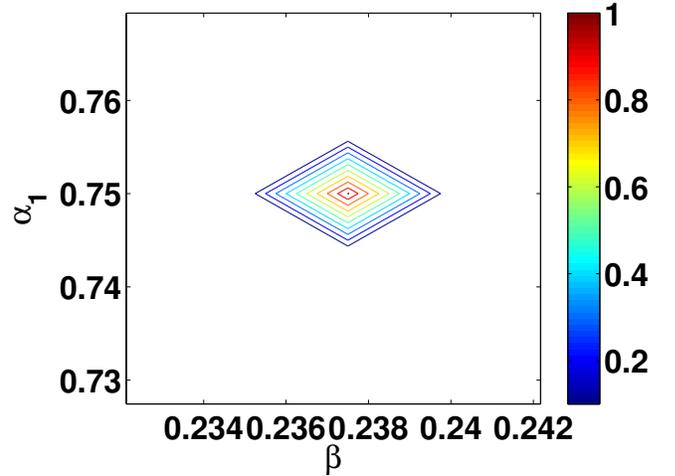


Fig. 5. Zoomed in Contour plot for the two parameter Duffing estimation problem

VI. RESULTS

Here, we consider $\alpha_1 = 0.75, \beta = 0.23$. A 3 dimensional plot of the probabilities is shown in Figure 4, and also a close up of the contour plot is presented in Figure 5.

It can clearly be seen that the anomaly has been well isolated and a single, sharp peak is obtained in the probability plot. The predicted values of α_1 lie between 0.74 and 0.76, while the values of β lie between 0.234 and 0.240, which is within a fairly high confidence value of the actual value selected for the test. In the forward problem, eight symbols are used, and the contours for each symbol are shown in the Figure 6

VII. SUMMARY AND CONCLUSIONS

This paper formulates and validates a real-time algorithm for simultaneous estimation of multiple parameters and anomaly pattern identification in the presence of (possibly slowly varying) anomalies. A central step in this kind of

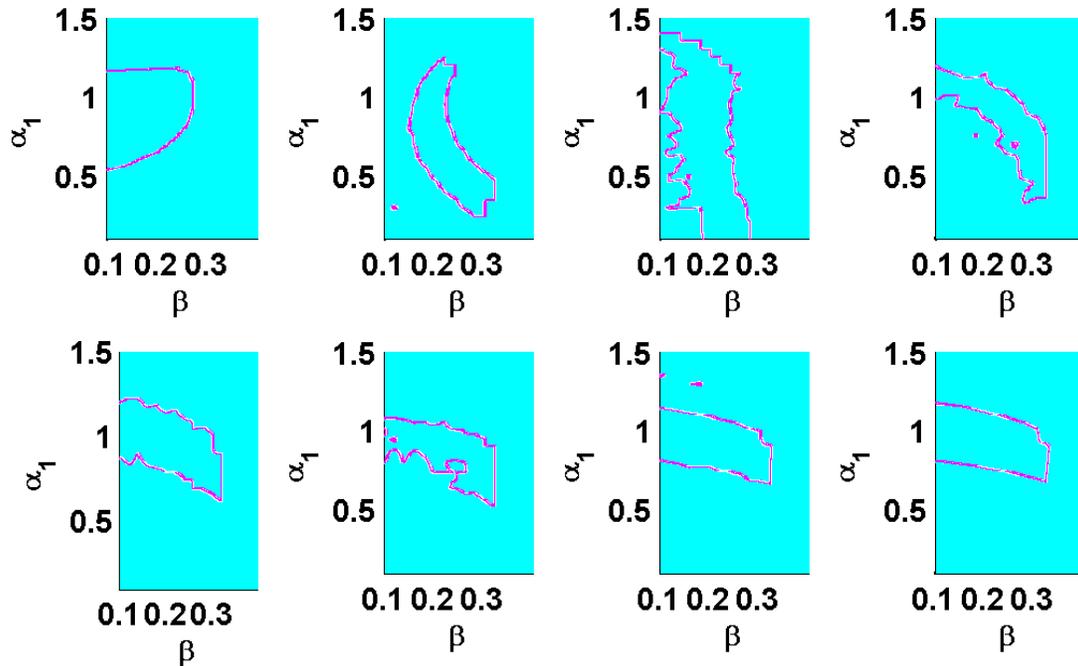


Fig. 6. Contour plot for each symbol for the two parameter Duffing estimation problem

identification methodology is based on analysis of the observed time-series data via conversion into a corresponding sequence of symbols. The algorithm is formulated in terms of Symbolic Dynamics and escort probabilities [5] in the setting of thermodynamic formalism.

This proposed algorithm has been tested for estimation of two slowly varying parameters in an active electronic system that is constructed in the classical Duffing equation setting. The proposed method is robust with respect to sensor noise, and computationally simple enough to be implemented in mobile platforms or even embedded within the sensor software system. The resulting compressed information can be conveniently transmitted over a sensor network and implemented efficiently with respect to memory requirements and processing needs.

REFERENCES

[1] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Processing*, vol. 84, no. 7, pp. 1115–1130, 2004.

[2] V. Rajagopalan and A. Ray, "Symbolic time series analysis via wavelet-based partitioning," *Signal Processing*, vol. 86, no. 11, pp. 3309–3320, 2006.

[3] S. Gupta, A. Ray, and E. Keller, "Symbolic time series analysis of ultrasonic data for early detection of fatigue damage," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 866–884, 2007.

[4] R. Duda, P. Hart, and D. Stork, "Pattern classification, 2/e," *John Wiley, New York*, pp. 915–930, 2001.

[5] C. Beck and F. Schögel, *Thermodynamics of Chaotic Systems: An Introduction*. Cambridge University Press, 1993.

[6] R. Badii and A. Politi, *Complexity, Hierarchical Structures and Scaling in Physics*. Cambridge, U.K: Cambridge University Press, 1997.

[7] S. C. V. Rajagopalan and A. Ray, "Estimation of slowly varying parameters in nonlinear systems via symbolic dynamic filtering," in Press, *Signal Processing*.

[8] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Boston, MA: Academic Press, 1998.

[9] J.M.T. Thompson and H.B. Stewart, "Nonlinear dynamics and chaos," *John Wiley, Chichester, United Kingdom*, 1986.

[10] S. Abe, "Geometry of escort distributions," *PHYSICAL REVIEW*, vol. 68, no. 3, pp. 031 101–+, Sept. 2003.