Signal Processing 89 (2009) 981–988

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Statistical estimation of multiple parameters via symbolic dynamic filtering ☆

Chinmay Rao, Kushal Mukherjee, Soumik Sarkar, Asok Ray *

Department of Mechanical Engineering, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 23 July 2008
Received in revised form
21 November 2008
Accepted 27 November 2008
Available online 16 December 2008

Keywords:
Parameter estimation
Symbolic dynamics
Time series analysis
Nonlinear dynamical systems

ABSTRACT

This paper addresses statistical estimation of multiple parameters that may vary simultaneously but slowly relative to the process response in nonlinear dynamical systems. The estimation algorithm is sensor-data-driven and is built upon the concept of symbolic dynamic filtering for real-time execution on limited-memory platforms, such as local nodes in a sensor network. In this approach, the behavior patterns of the dynamical system are compactly generated as quasi-stationary probability vectors associated with the finite-state automata for symbolic dynamic representation. The estimation algorithm is validated on nonlinear electronic circuits that represent externally excited Duffing and unforced van der Pol systems. Confidence intervals are obtained for statistical estimation of two parameters in each of the systems.

1. Introduction

Recent literature has reported various methods for estimation of multiple parameters, such as those based on joint state estimation [1], parity equations [2], generalized likelihood ratio [3], Karhunen–Loéve and Galerkin multiple shooting [4], similarity measures [5], and orthogonal Haar transform [6]. An application of parameter estimation is the detection and mitigation of evolving faults in interconnected dynamical systems [7]. Often evolution of gradual deviations from the nominal behavior in individual components of such systems may lead to cascaded faults because of strong input–output and feedback interconnections between the system components, and may eventually cause catastrophic failures and forced shutdown of the entire system. In such a scenario, the problem of degradation monitoring of the system reduces to simultaneous estimation of several slowly varying critical parameters.

Many conventional methods of parameter estimation are model-based and they are often inadequate for human-engineered complex systems due to unavailability of a reliable model of the process dynamics. To alleviate this problem, data-driven parameter estimation methods have been formulated in the setting of hidden Markov models (*HMM*) [8]. One such method is symbolic dynamic filtering (*SDF*) [9,10] that is based on the concept of symbolic time series analysis (*STSA*) [11]; *SDF* belongs to the class of data-driven statistical pattern recognition and enables compression of information into pattern vectors of low dimension for real-time execution on limited-memory platforms, such as small microprocessors in a sensor network. In a recent publication [12], performance of *SDF* has been shown to be superior to that of several pattern classification techniques such as principal component analysis (*PCA*), artificial neural networks (*ANN*), kernel regression analysis (*KRA*), particle filtering (*PF*) and unscented Kalman filtering (*UKF*), in terms of early

☆ This work has been supported in part by the US Army Research Office (ARO) under Grant no. W911NF-07-1-0376 and by NASA under Cooperative Agreement no. NNX07AK49A. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

* Corresponding author. Tel.: +1 814 865 6377.
E-mail addresses: crr164@psu.edu (C. Rao), kum162@psu.edu (K. Mukherjee), szs200@psu.edu (S. Sarkar), axr2@psu.edu (A. Ray).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2008.11.018

detection of changes, computation speed and memory requirements.

As an extension of the parameter estimation method of Tang et al. [13], which is also based on *STSA*, Piccardi [14] proposed multiple-parameter estimation in chaotic systems; although the symbolic analysis was performed on a probabilistic finite-state model, the parameter vector was estimated by (genetic algorithm) optimization in a deterministic setting. The present paper, which is built upon the concept of *SDF*, proposes an alternative approach to estimation of multiple parameters as described below.

The framework of *SDF* includes preprocessing of time series data by time–frequency analysis (e.g., wavelet transform [15] and Hilbert transform [16,17]). The transformed data set is partitioned using the maximum entropy principle [18] to generate the symbol sequences from the transformed data set without any significant loss of information. Subsequently, statistical patterns of the evolving system dynamics are identified from these symbol sequences through construction of probabilistic finite-state automata (*PFSA*). An additional advantage of transform space-based partitioning is reduction of spurious noise in the data set from which the *PFSA* is constructed; this feature provides additional robustness to *SDF* as discussed in [18]. The state probability vectors that are derived from the respective state transition probability matrices of *PFSA* serve as behavioral patterns of the evolving dynamical system under nominal and off-nominal conditions.

Parameter estimation algorithms, based on *SDF*, have been experimentally validated for real-time execution in different applications, such as degradation monitoring in electronic circuits [12] and fatigue damage monitoring in polycrystalline alloys [19]. While these applications of *SDF* have focused on estimation of only a single parameter, the work reported here addresses statistical estimation of multiple parameters. Specifically, this paper is an extension of the earlier work [20] on single-parameter estimation to estimation of multiple parameters that may vary simultaneously. The resulting algorithms are validated on the same test apparatus as [20] for the following electronic systems:

(1) Externally excited Duffing system [21]:

$$\frac{d^2x(t)}{dt^2}+\beta\frac{dx}{dt}+\alpha_1 x(t)+x^3(t)=A\cos(\omega_e t) \quad (1)$$

where the amplitude $A=22.0$, excitation frequency $\omega_e=5.0$, and nominal values of the parameters, to be estimated, are $\alpha_1=1.0$ and $\beta=0.1$.

(2) Unforced van der Pol system [22]:

$$\frac{d^2x(t)}{dt^2}-\mu(1-x^2(t))\frac{dx(t)}{dt}+\omega^2 x(t)=0 \quad (2)$$

where nominal values of the parameters, to be estimated, are $\mu=1.0$ and $\omega=1.0$.

While the parameter estimation algorithm is tested on an experimental apparatus, a system model is generally used for the purpose of training. Therefore, model reliability or statistic of the modeling error is crucial for robustness of the algorithm and should be known *a priori*. However, this issue is not within the scope of this paper as both training and testing are carried out on similar experimental devices.

2. Review of SDF and single-parameter estimation

This section succinctly reviews the theory of *SDF* [9] and explains the underlying concept of single-parameter estimation [20] in the *SDF* framework.

Extraction of statistical behavior patterns from time series data is posed as a two-scale problem. The *fast scale* is related to response time of the process dynamics. Over the span of data acquisition, dynamic behavior of the system is assumed to remain invariant, i.e., the process is quasi-stationary at the fast scale. In other words, variations in the statistical behavior of the dynamical system are assumed to be negligible on the fast scale. The *slow scale* is related to the time span over which deviations (e.g., parametric changes) may occur and exhibit non-stationary dynamics. The parameters are estimated based on the information generated by *SDF* of the data collected over the fast scale at a slow scale epoch. This method is also applicable to estimation of slowly varying parameters. The rationale is that, since the parameters vary slowly, they are treated as invariants at a given slow scale epoch; accordingly, the fast-scale statistical behavior of the dynamical system may change at different slow scale epochs (that are simply referred to as epochs in the sequel).

2.1. Forward problem in the symbolic dynamic setting

This subsection summarizes the *forward problem* for detection of deviation patterns in the *SDF* setting:

(1) *Time series data acquisition on the fast scale from sensors and/or analytical measurements* (i.e., *outputs of a physics-based or an empirical model*). Data sets are collected at the parameter values as a set $\{s^0,s^1,\ldots,s^k,\ldots\}$, where s^k denotes the value of the parameter at the epoch k.

(2) *Generation of wavelet transform coefficients with an appropriate choice of the wavelet basis and scales*. The wavelet transform largely alleviates the difficulties of phase-space partitioning and is particularly effective with noisy data from high-dimensional dynamical systems.

(3) *Maximum entropy partitioning of the wavelet space at a reference condition*. Each segment of the partitioning is assigned a particular symbol from the symbol alphabet Σ. This step enables transformation of time series data from the continuous domain to the symbolic domain [23].

(4) *Construction of a probabilistic finite-state automaton* (*PFSA*) *at the reference condition*. The structure of the finite-state machine is fixed for subsequent parameter values until a new reference is selected.

(5) *Computation of the reference pattern vector* $\mathbf{p}(s^0)$ *whose elements represent state occupation probabilities of the PFSA at the reference condition*. Such a pattern vector is recursively computed as an approximation of the

natural invariant density of the dynamical system, which is a fixed point of the local Perron–Frobenius operator [24]. Thus, $\mathbf{p}(s^0) \equiv [p_1(s^0)\ p_2(s^0)\ \cdots\ p_{|\Sigma|}(s^0)]$, where $|\Sigma|$ is the number of states in the *PFSA*.

(6) *Time series data acquisition on the fast scale at subsequent parameter values, and their conversion to respective symbolic sequences based on the reference partitioning at the reference value.*

(7) *Generation of the pattern vectors,* $\mathbf{p}(s^1), \mathbf{p}(s^2), \ldots, \mathbf{p}(s^k), \ldots$ *at parameter values,* $s^1, s^2, \ldots, s^k, \ldots$ *from the respective symbolic sequences using the state machine constructed at nominal parameter value* s^0. Thus, $\mathbf{p}(s^k) \equiv [p_1(s^k)\ p_2(s^k)\ \cdots\ p_{|\Sigma|}(s^k)]$, where $|\Sigma|$ is the number of states in the *PFSA*. (Note that only $(|\Sigma| - 1)$ out of the $|\Sigma|$ elements of $\mathbf{p}(s^k)$ are linearly independent because $\mathbf{p}(s^k)$ is sum-normalized to unity.) The structure of the *PFSA* at all epochs is identical in the *SDF* framework, while the pattern vectors $\mathbf{p}(s^k)$ are possibly different at different parameter values s^k.

(8) *Computation of deviation measures*: Evolving deviation measures $\mathscr{M}(s^1), \mathscr{M}(s^2), \ldots, \mathscr{M}(s^k), \ldots$ at parameter values, $s^1, s^2, \ldots, s^k, \ldots$, are computed with respect to the nominal condition at s^0, by selecting an appropriate distance function $d(\bullet, \bullet)$ (e.g., the standard Euclidean norm) such that

$$\mathscr{M}(s^k) \triangleq d(\mathbf{p}(s^k), \mathbf{p}(s^0)) \tag{3}$$

2.2. Inverse problem of single-parameter estimation

This subsection focuses on the inverse problem of single-parameter estimation based on computed values of the deviation measure in the forward problem. The parameter to be estimated is treated as a random variable at each epoch, for which the deviation measure is an observable. 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 sets of time series data during each experiment, with the objective of estimating the unknown parameter. The steps for the statistical identification of the system parameter from the measured value of deviation measure are delineated below:

(1) *Upon generation of deviation measure profiles in the forward problem, a statistical relationship is identified between deviation measure and the parameter associated with the deviation.* In particular, probability distributions of the parameter are obtained for various values of the deviation measure. Then, statistical tests are performed to determine goodness-of-fit of the distributions. For example, mean and variance associated with a two-parameter distribution provide adequate statistical information on the bounds and confidence levels of the estimated parameter.

(2) *Data acquisition on the fast scale at an unknown parameter value.* Time series data are collected (in the fast scale) under operating conditions similar to those in Step 1 of the forward problem. Data are analyzed to generate pattern vectors as described in the forward problem. The deviation measure $\mathscr{M}^{\text{test}}$ at parameter value s^{test} is then calculated by quantifying the deviation of the current pattern vector $\mathbf{p}^{\text{test}}$ from the nominal pattern vector $\mathbf{p}(s^1)$.

(3) *Parameter estimation from generated statistics of deviation profile.* The estimated value of the parameter and its confidence interval are obtained based on the computed deviation measure and the probability distribution derived in Step 1 of the inverse problem.

In the above procedure, the range of the computed deviation measure profile is discretized into finitely many levels. A statistical distribution is hypothesized for determining spread of the parameter and goodness-of-fit of the hypothesized distribution that is assessed with χ^2 and Kolmogorov–Smirnov tests [25].

3. Framework of multi-parameter estimation

In general, extension of single-parameter estimation [20] to multiple-parameter extension is not a straightforward task as explained below.

Let us consider the Duffing system in Eq. (1), where the parameters to be estimated are chosen as α_1 and β; and the deviation measure $\mathscr{M}$ (see Eq. (3)) is obtained for the parameter pair $\underline{s} \triangleq (s_1, s_2) = (\alpha_1, \beta)$. Fig. 1 shows a plot with α_1 on the x-axis, β on the y-axis and the contours of the deviation measure $\mathscr{M}$. Each contour is constructed by joining points with the same value of deviation measure $\mathscr{M}$; this is indicated by the gray scale (color) corresponding to the vertical bar on the right hand side of the plot. Values of deviation measure $\mathscr{M}$ are chosen in steps of 0.1 and a plane parallel to the x–y axis is constructed at these values of $\mathscr{M}$ to join points of equal values of deviation measure. As the system deviates in either direction from the nominal condition of $\alpha_1 = 1.0$ and $\beta = 0.1$, the deviation measure $\mathscr{M}$ increases until bifurcation occurs

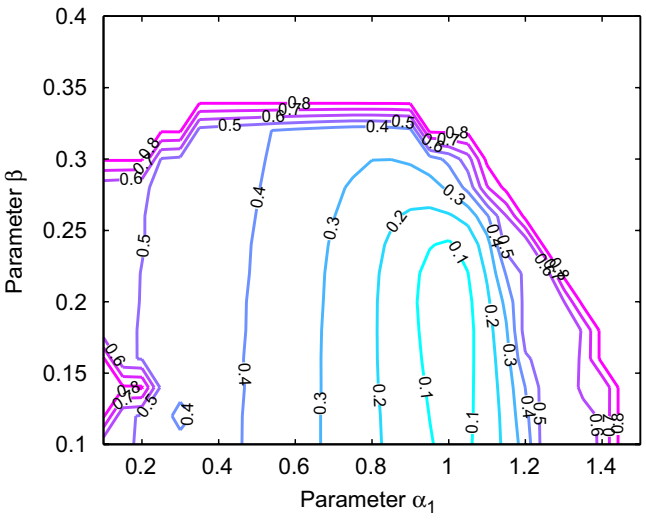

Fig. 1. Contour plot of the deviation measure $\mathscr{M}$.

(e.g., $\alpha_1 = 1.0$ and $\beta \approx 0.32$). It is obvious that the inverse image of a singleton set of $\mathscr{M}$ may contain infinitely many combinations of α_1 and β. Hence, the information on $\mathscr{M}$ alone is insufficient for uniquely identifying the parameters α_1 and β that characterize the system. It demonstrates that non-uniqueness of estimation could occur if a scalar-valued function is chosen as the cost functional for optimized estimation of multiple parameters. This problem is resolved by considering the individual elements of the frequency probability vector for statistical estimation of the parameters as explained below.

If the estimation of multiple parameters is set as an optimization problem with deviation measure $\mathscr{M}$ being the cost functional, then non-convexity may arise due to existence of contours; this situation could occur even if the range of optimization is narrow. Therefore, instead of relying on the deviation measure for parameter estimation $\mathscr{M}(s^k)$, as it was done in [13,14,20], variations in the individual elements of $\mathbf{p}(s^k) \triangleq [p_1(s^k), p_2(s^k), \ldots, p_{|\Sigma|}(s^k)]$ are used in this paper. That is, the parameter estimation problem is reduced to identification of contours for $(|\Sigma| - 1)$ independent elements of the state probability vector $\mathbf{p}(s^k)$. The information derived from these $(|\Sigma| - 1)$ independent contours would yield a statistical estimate of the parameter vector $\underline{s}^k$. This approach circumvents the aforementioned non-convexity problem.

For a given parameter pair of the Duffing system having values $s^k = (\alpha_1^k, \beta^k)$, the pattern vectors are generated as $\mathbf{p}(s^k) \equiv [p_1(s^k)\ p_2(s^k)\ \cdots\ p_{|\Sigma|}(s^k)]$. Having $|\Sigma| = 8$, eight plates in Fig. 2 shows contours for each of the eight elements of $\mathbf{p}(s^k)$ for a given value of $(\alpha_1^k = 0.75,\ \beta^k = 0.23)$. The following two subsections describe a method that makes use of the ensemble of information in different contours to arrive at a more precise estimation of the parameters.

3.1. Construction of the multi-parameter algorithm

Let $\mathscr{S}$ denote the collection of (finitely many) points in the n-dimensional parameter space, where the positive integer n is the number of parameters that are to be estimated. That is, $\mathscr{S} = \{\underline{s}^1, \underline{s}^2, \ldots, \underline{s}^{|\mathscr{S}|}\}$, on which the training process is executed. Let Ω be the convex hull of $\mathscr{S}$, which represents the range over which the parameters take values. It is noted that Ω is a convex and compact subset of the separable space $\mathbb{R}^n$.

Let each element $\underline{s}^k \triangleq (s_1^k, s_2^k, \ldots, s_n^k)$ represent a particular set of parameters. For the Duffing system in Eq. (1),

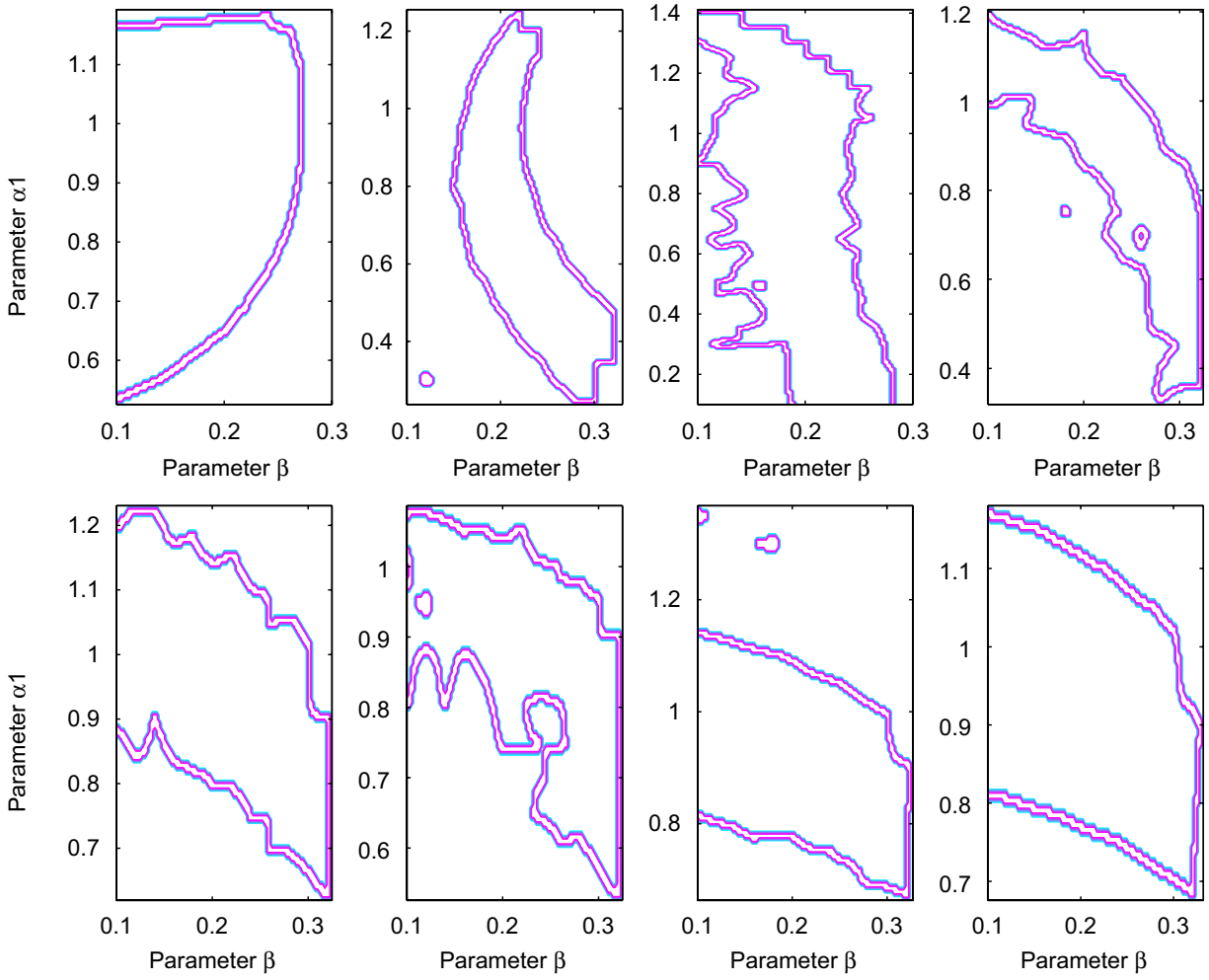

Fig. 2. Contour plots of each element of the probability vector $\mathbf{p}(s^k)$ for a typical test case in the Duffing system.

the set $\mathscr{S}$ consists of different values of parameters α_1 and β in the range where the experiments have been conducted; for example, a permissible value of $\underline{s}$ is $(\alpha_1,\ \beta) = (0.3,\ 0.4)$. Given an experimental time series data set Υ, the problem at hand is to identify the conditional probability density $f(\underline{s}\,|\Upsilon)$, where $\underline{s} \in \Omega$. The procedure of multi-parameter estimation consists of the forward problem and the inverse problem that are analogous to, but much more involved than, the single-parameter estimation procedure described in Section 2.

3.1.1. Forward problem/training

For the forward problem, sets of time series data are generated by experimental runs at parameter values $\underline{s}^k,\ \forall k = 1,2,\ldots,|\mathscr{S}|$. A symbolic dynamic filter is constructed to analyze each data sequence as outlined in Section 2. For $|\Sigma|$ being the number of automaton states, the n-dimensional pattern vector $\mathbf{p}(\underline{s}^k)$ is generated for every $\underline{s}^k \in \mathscr{S}$. The procedure for data acquisition and storage for statistical analysis is described below.

For each $\underline{s}^k \in \mathscr{S}$, L different samples of the random parameter vector were collected, which were realized as identically manufactured but different electronic cards in experimental apparatus [20]. This implies that $L \times |\mathscr{S}|$ experiments need to be conducted on the apparatus. In this paper, there were $L = 40$ different realizations of the experimental apparatus. For each of Duffing and van der Pol experiments, $|\mathscr{S}| = 50$ and hence the total number of experiments was $40 \times 50 = 2000$.

Let the elements $p_j(s^k)$, $j = 1,2,\ldots,|\Sigma|$ of the state probability vector $\mathbf{p}(s^k)$, $k = 1,2,\ldots,|\mathscr{S}|$ be modeled as a random variable $q_j(s^k)$ that is constructed from the ensemble of data points. The resulting random vector is obtained as

$$\mathbf{q}(s^k) \equiv [q_1(s^k)\ \ q_2(s^k)\ \ \cdots\ \ q_{|\Sigma|}(s^k)] \tag{4}$$

where $q_j(s^k) \sim \mathscr{N}[m_j(s^k), \sigma_j^2(s^k)]$, i.e., $q_j(s^k)$ is modeled to be Gaussian with mean $m_j(s^k)$ and variance $\sigma_j^2(s^k)$, as explained below from the perspectives of state machine construction in the *SDF* setting. The equation for the modeled distribution is given as

$$f_{q_j|\mathscr{S}}(p_j|s^k) = \frac{1}{\sqrt{2\pi\sigma_j^2(s^k)}} \exp\left(-\frac{(p_j - m_j(s^k))^2}{2\sigma_j^2(s^k)}\right) \tag{5}$$

The underlying dynamical system is modeled as an irreducible Markov process via *SDF*, where the state probability vector is the sum-normalized eigenvector of the state transition matrix corresponding to the unique unity eigenvalue. Hence, no element in the state probability vector is either 0 or equal to 1. However, due to process noise and sensor noise, the random variable $q_j(s^k)$ fluctuates around its mean $m_j(s^k)$. While analyzing the experimental data, the standard deviation $\sigma_j(s^k)$ of the random variables $q_j(s^k)$ was found to be very small compared to its expected value $m_j(s^k)$, i.e., the ratio $\sigma_j(s^k)/m_j(s^k) \ll 1\ \forall k = 1,2,\ldots,|\mathscr{S}|\ \forall j = 1,2,\ldots,|\Sigma|$. Therefore, a parametric or non-parametric two-sided unimodal distribution should be adequate to model the random variable $q_j(s^k)$. The choice of Gaussian distribution for q_j would facilitate estimation of the statistical parameters and involve only second order statistics. This assumption has been validated by using the χ^2 and Kolmogorov–Smirnov tests for goodness-of-fit [25] of each q_j for Gaussian distribution.

Remark 3.1. The random variables $q_j(s^k)$ must satisfy the following two conditions:

- Positivity, i.e., $q_j(s^k) > 0\ \forall \underline{s} \in \mathscr{S}\ \forall j = 1,2,\ldots,|\Sigma|$. This is made possible by truncating the far end of the Gaussian distribution tail on the left side. The goodness-of-fit of the distribution as Gaussian still remains valid at a very high significance level.
- Unity sum of the state probabilities, i.e., $\sum_{j=1}^{|\Sigma|} q_j(s^k) = 1\ \forall \underline{s} \in \mathscr{S}$. This is achieved by sum-normalization.

Remark 3.2. The automaton states are analogous to energy states in statistical mechanics of ideal gases [26]. This fact is used for formulating the inverse problem as explained below.

3.1.2. Inverse problem/testing

Let time series data be generated from a new test on the experimental apparatus. The task at hand is to identify, from this data set, the unknown parameter vector $\underline{s} \in \Omega$; however, it is possible that $\underline{s} \notin \mathscr{S}$. The data are analyzed using the same symbolic dynamic filter constructed in the forward/training problem (see Section 3.1.1), and the resulting probability vector $\mathbf{p} \equiv [p_1 \cdots p_{|\Sigma|}]$ is a realization of a random vector $\mathbf{q} \equiv [q_1 \cdots q_{|\Sigma|}]$. The density function $f_{\Omega|q}(\underline{s}\,|\mathbf{p})$ is obtained as

$$f_{\Omega|\mathbf{q}}(\underline{s}\,|\mathbf{p}) = \frac{f_{\mathbf{q}|\Omega}(\mathbf{p}|\,\underline{s}) f_\Omega(\underline{s})}{f_{\mathbf{q}}(\mathbf{p})} = \frac{f_{\mathbf{q}|\Omega}(\mathbf{p}|\,\underline{s}) f_\Omega(\underline{s})}{\int_\Omega f_{\mathbf{q}|\Omega}(\mathbf{p}|\,\tilde{\underline{s}}) f_\Omega(\tilde{\underline{s}}) d\,\tilde{\underline{s}}} \tag{6}$$

In the absence of *a priori* information, an assumption is made that all operating conditions are equally likely, i.e., $f_\Omega(\underline{s}) = f_\Omega(\tilde{\underline{s}})\ \forall\ \underline{s},\ \tilde{\underline{s}} \in \Omega$. With this assumption of uniform probability, Eq. (6) reduces to

$$f_{\Omega|\mathbf{q}}(\underline{s}\,|\mathbf{p}) = \frac{f_{\mathbf{q}|\Omega}(\mathbf{p}|\,\underline{s})}{\int_\Omega f_{\mathbf{q}|\Omega}(\mathbf{p}|\,\tilde{\underline{s}})\,\mathrm{d}\,\tilde{\underline{s}}} \tag{7}$$

It is noted that accuracy of the above distribution would be improved if the actual prior mapping, i.e., $f_\Omega(\tilde{\underline{s}})$ is known. The integral in the denominator of Eq. (7) is approximated by a Reimann sum as

$$f_{\Omega|\mathbf{q}}(\underline{s}\,|\mathbf{p}) \approx \kappa \frac{f_{\mathbf{q}|\Omega}(\mathbf{p}|\,\underline{s})}{\sum_{\mathscr{S}} f_{\mathbf{q}|\Omega}(\mathbf{p}|\,\tilde{\underline{s}})} \tag{8}$$

where κ is a constant. This approximation converges to the exact solution as the training set $\mathscr{S}$ approaches a (countable) dense subset of $\Omega \subset \mathbb{R}^n$.

The density function in Eq. (8) is now sampled at the points $\underline{s}^k$ in the training set $\mathscr{S}$ to construct the following sampled density to yield

$$f_{\Omega|\mathbf{q}}(\underline{s}\,|\mathbf{p})\big|_{\underline{s}=\underline{s}^k} \approx \kappa \frac{f_{\mathbf{q}|\Omega}(\mathbf{p}|\underline{s}^k)}{\sum_{\tilde{\underline{s}} \in \mathscr{S}} f_{\mathbf{q}|\Omega}(\mathbf{p}|\,\tilde{\underline{s}})} \quad \forall \underline{s}^k \in \mathscr{S} \tag{9}$$

Furthermore, it is observed from experimental data that fluctuations of p_i are uncorrelated with those of p_j for all $i \neq j$, where $i,j = 1,2,\ldots,|\Sigma|$. Therefore, the joint density function of the Gaussian random vector $\mathbf{p}$ is reduced to the

product of individual Gaussian distributions of the random variables p_j. That is,

$$f_{\Omega|\mathbf{q}}(\underline{s}\,|\mathbf{p})|_{\underline{s}=\underline{s}^k} \approx \kappa \frac{\prod_{j=1}^{|\Sigma|-1} f_{q_j|\Omega}(p_j|\underline{s}^k)}{\sum_{\underline{\tilde{s}}\in\mathscr{S}} \prod_{j=1}^{|\Sigma|-1} f_{q_j|\Omega}(p_j|\underline{\tilde{s}}^k)} \tag{10}$$

The density functions in the numerator and denominator of Eq. (10) are obtained from Eqs. (9) and (5), which were determined in the training phase. A most likely estimate of the parameter vector $\underline{s}$ is obtained from the probabilistic map in Eq. (10). It should be noted that the nature of the density function $f_{\Omega|\mathbf{q}}(\underline{s}^k|\mathbf{p})$ does not depend on the constant κ.

The probability mass functions are obtained by evaluating the probability density function in Eq. (9) at points $\underline{s}^k \in \mathscr{S}$.

$$P(\underline{s}^k|\mathbf{p}) \triangleq \frac{f_{\Omega|\mathbf{q}}(\underline{s}^k|\mathbf{p})}{\sum_{j=1}^{|\mathscr{S}|} f_{\Omega|\mathbf{q}}(\underline{s}^j|\mathbf{p})} \approx \frac{f_{\mathbf{q}|\Omega}(\mathbf{p}|\underline{s}^k)}{\sum_{j=1}^{|\mathscr{S}|} f_{\mathbf{q}|\Omega}(\mathbf{p}|\underline{s}^j)} \tag{11}$$

Substitution of Eqs. (5) and (10) in Eq. (11) yields

$$P(\underline{s}^k|\mathbf{p}) \approx \frac{\prod_{j=1}^{|\Sigma|-1} \frac{1}{\sqrt{2\pi\sigma_j^2(\underline{s}^k)}} \exp\left(-\frac{(p_j - m_j(\underline{s}^k))^2}{2\sigma_j^2(\underline{s}^k)}\right)}{\sum_{l=1}^{|\mathscr{S}|} \prod_{j=1}^{|\Sigma|-1} \frac{1}{\sqrt{2\pi\sigma_j^2(\underline{s}^l)}} \exp\left(-\frac{(p_j - m_j(\underline{s}^l))^2}{2\sigma_j^2(\underline{s}^l)}\right)} \tag{12}$$

where the probability vector $\mathbf{p} \equiv [p_1 \cdots p_{|\Sigma|}]$ is calculated from the observed time series data; and the remaining parameters are already evaluated in the training phase.

Estimated mean $\hat{\underline{s}}$ and estimated covariance matrix $\hat{C}_{\underline{s}}$ of the parameter set $\underline{s}$, where $\underline{s}$ is in the column vector form, are obtained directly from Eq. (12) as

$$\hat{\underline{s}}(\mathbf{p}) \triangleq \sum_{k=1}^{|\mathscr{S}|} \underline{s}^k P(\underline{s}^k|\mathbf{p}) \tag{13}$$

$$\hat{C}_{\underline{s}}(\mathbf{p}) \triangleq \sum_{k=1}^{|\mathscr{S}|} (\underline{s}^k - \hat{\underline{s}}(\mathbf{p}))P(\underline{s}^k|\mathbf{p})(\underline{s}^k - \hat{\underline{s}}(\mathbf{p}))^{\mathrm{T}} \tag{14}$$

Since the statistical information is available in the form of probability mass functions, the third and higher moments of the parameter vector can be estimated in a similar way; however, third and higher moments are redundant because the inherent distribution is assumed to have a Gaussian structure that carries full statistical information in the first two moments.

4. Test results and discussion

This section presents the test results of multiple-parameter estimation on two electronic circuits, namely the externally excited Duffing system [21] and the unforced van der Pol system [22], on the test apparatus described in a previous publication [20].

4.1. Test results on the duffing system

This subsection analyzes and presents the experimental results for multiple-parameter estimation in the Duffing system described by Eq. (1). For the forward problem/training (see Section 3.1), training data sets were generated with α_1 ranging from 0.10 to 1.50 in steps of 0.05, and β ranging from 0.10 to 0.40 in steps of 0.02, and; the nominal condition was chosen as $\alpha_1 = 1.0$ and $\beta = 0.1$; and an *SDF* was constructed with the number of states in the automaton $|\Sigma| = 8$. This information on time series data was then fed into the *SDF* to compute the components p_j of pattern vectors $\mathbf{p}$ at different values of the parameter pair $(\alpha_1,\ \beta)$. As the dynamics of the Duffing system changed due to variations in the parameters α_1 and β, the statistics of the symbol sequences were altered and so were the pattern vectors.

For the inverse problem/testing (see Section 3.1.2), experiments were conducted at the assigned values of the parameters that were different from those in the forward problem of *SDF* but within the range of α_1 and β where the training was conducted. The components p_j of pattern vectors $\mathbf{p}$ at different values of the parameter pair $(\alpha_1,\ \beta)$ were computed from the data sets that were generated with these assigned values of parameters. For a typical test at $\alpha_1 = 0.75$ and $\beta = 0.23$, the 3-dimensional plot in Fig. 3 shows the bivariate probability distribution, followed by a close-up view of the contour plots in Fig. 4. The parameter pair $(\alpha_1,\ \beta)$ is crisply identified by a single, sharp spike in the probability distribution plot of Fig. 3, where the estimates $\hat{\alpha}_1$ and $\hat{\beta}$ lie in the ranges of (0.745, 0.755) and (0.235, 0.240), respectively, as seen in Fig. 4. Table 1 shows the results for mean, standard deviation, and confidence intervals of the parameter estimates, $\hat{\alpha}_1$ and $\hat{\beta}$ for test runs with four different pairs of α_1^{test} and β^{test} that do not belong to the set $\mathscr{S}$ of training data. It is seen that the estimated mean values of both α_1 and β are orders of magnitude greater than their respective standard deviations $\hat{\sigma}_{\alpha_1} \triangleq \sqrt{\hat{C}_{\alpha_1\alpha_1}}$ and $\hat{\sigma}_{\beta} \triangleq \sqrt{\hat{C}_{\beta\beta}}$. This observation suggests that the estimates are relatively close to the true values of the parameters. It is also seen in Table 1 that the correlation coefficient $\hat{C}_{\alpha_1\beta}/\hat{\sigma}_{\alpha_1}\hat{\sigma}_{\beta}$ is a positive fraction, which implies that the parameters α_1 and β are positively correlated. The rationale for this correlation is that variations even in a single component of a dynamical system may cause simultaneous variations in several parameters of its

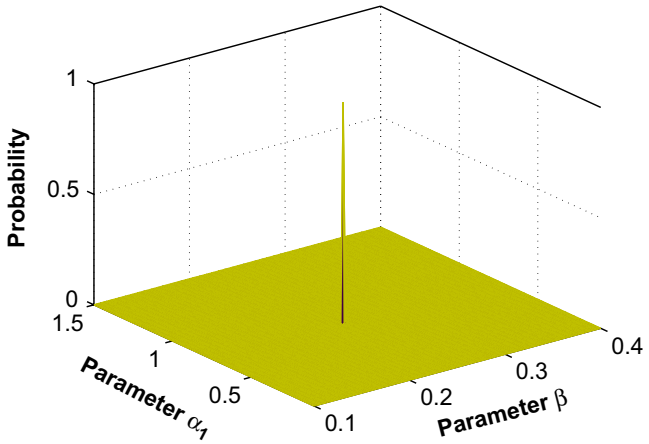

Fig. 3. Joint probability distribution of the parameter pair α_1 and β.

governing equations. In the Duffing system, usage of different but identically manufactured electronic cards caused simultaneous variations of both parameters α_1 and β in Eq. (1).

4.2. Results on van der Pol System

This subsection analyzes and presents experimental results for multiple-parameter estimation in the van der Pol system described by Eq. (2). For the forward problem/training (see Section 3.1), training data sets were generated with both parameters μ and ω ranging from 0.5 to 4.0 in increments of 0.5; and an *SDF* was constructed with the number of states in the automaton $|\Sigma| = 8$. This information on time series data was then fed into the *SDF* to compute pattern vectors **p** and deviation measures $\mathscr{M}$ at different values of the parameter pair (μ, ω). As the dynamics of the van der Pol system changed due to variations in the parameters μ and ω, the statistics of the symbol sequences were altered and so were the pattern vectors. However, unlike the Duffing system, no abrupt change (e.g., bifurcation) in the dynamic behavior was observed as μ and ω were varied from the nominal condition.

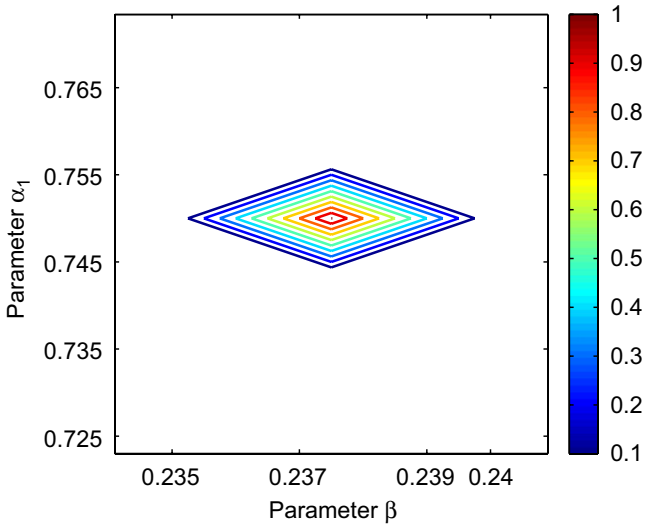

Fig. 4. Zoomed-in contour plots of the parameter pair α_1 and β.

For the inverse problem/testing (see Section 3.1.2), experiments were conducted at the assigned values of the parameters that were different from those in the forward problem of *SDF*, i.e., they do not belong to the set $\mathscr{S}$ of training data. Pattern vectors **p** and the associated deviation measures $\mathscr{M}$ were estimated from of the data sets generated with these assigned values of parameters. Table 2 shows the results for mean, standard deviation, and confidence intervals of the parameter estimates, $\hat{\mu}$ and $\hat{\omega}$ for test runs with four different pairs of μ^{test} and ω^{test}. It is seen that the estimated mean values of both μ and ω are orders of magnitude greater than their respective standard deviations $\hat{\sigma}_\mu \triangleq \sqrt{\hat{C}_{\mu\mu}}$ and $\hat{\sigma}_\omega \triangleq \sqrt{\hat{C}_{\omega\omega}}$. This observation suggests that the estimates are relatively close to the true values of the parameters. It is also seen in Table 2 that the correlation coefficient $\hat{C}_{\mu\omega}/\hat{\sigma}_\mu\hat{\sigma}_\omega$ is close to 0, implying that the parameters μ and ω are very weakly correlated.

5. Summary, conclusion, and future work

This paper presents an application of *SDF* [9] for multiple-parameter estimation in nonlinear dynamical

Table 1
Predicted values of $(\hat{\alpha_1}, \hat{\beta})$ and confidence intervals for the Duffing equation.

Estimates										
Parameter α_1			$\hat{C}_{\alpha_1\beta}$	Parameter β			95% confidence interval		90% confidence interval	
α_1^{test}	$\hat{\alpha}_1$	$\hat{\sigma}_{\alpha_1}$		β^{test}	$\hat{\beta}$	$\hat{\sigma}_\beta$	$(\alpha_{1_{\min}}, \alpha_{1_{\max}})$	$(\beta_{\min}, \beta_{\max})$	$(\alpha_{1_{\min}}, \alpha_{1_{\max}})$	$(\beta_{\min}, \beta_{\max})$
0.30	0.30	8.4e−4	2.67e−7	0.10	0.10	4.0e−4	(0.30, 0.30)	(0.30, 0.30)	(0.10, 0.10)	(0.10, 0.10)
0.45	0.46	0.015	5.15e−4	0.20	0.20	0.057	(0.45, 0.46)	(0.20, 0.20)	(0.45, 0.46)	(0.20, 0.20)
0.15	0.15	3.3e−3	2.465e−5	0.14	0.14	8.0e−3	(0.15, 0.15)	(0.14, 0.14)	(0.15, 0.15)	(0.14, 0.14)
0.65	0.65	3.0e−3	8.16e−6	0.35	0.36	7.0e−3	(0.65, 0.65)	(0.36, 0.36)	(0.65, 0.65)	(0.36, 0.36)

Table 2
Predicted values of $(\hat{\mu}, \hat{\omega})$ and confidence intervals for the van der Pol equation.

Estimates										
Parameter μ			$\hat{C}_{\mu\omega}$	Parameter ω			95% confidence interval		90% confidence interval	
μ^{test}	$\hat{\mu}$	$\hat{\sigma}_\mu$		ω^{test}	$\hat{\omega}$	$\hat{\sigma}_\omega$	$(\mu_{\min}, \mu_{\max})$	$(\omega_{\min}, \omega_{\max})$	$(\mu_{\min}, \mu_{\max})$	$(\omega_{\min}, \omega_{\max})$
2.50	2.50	0.006	−6.7e−6	1.00	1.00	0.028	(2.49, 2.51)	(0.99, 1.01)	(2.49, 2.50)	(0.99, 1.00)
3.30	3.32	0.059	3.0e−4	3.40	3.42	0.035	(3.29, 3.35)	(3.39, 3.43)	(3.30, 3.33)	(3.40, 3.42)
4.00	3.99	0.069	1.1e−5	2.50	2.49	0.087	(3.98, 4.02)	(2.46, 2.52)	(3.99, 4.00)	(2.48, 2.50)
3.50	3.51	0.016	2.5e−4	4.00	3.99	0.220	(3.48, 3.52)	(3.75, 4.12)	(3.49, 3.51)	(3.89, 4.08)

systems. The reported work is an extension of earlier work [20] on single-parameter estimation and is validated on an experimental apparatus for electronic circuits of externally excited Duffing and unforced van der Pol systems.

The proposed parameter estimation tool is sensor-data-driven and is apparently suitable for applications such as early detection of parametric faults for prognosis of catastrophic failures in human-engineered systems. A previous publication [12] has shown that the training process of the *SDF*-based parameter estimation method is significantly less time-consuming than those of mutilayer-perceptron and radial-basis-function neural networks. This is so because the underlying algorithm of *SDF* makes use of a stopping rule [12] to limit the length of the time series and then compresses the pertinent information into pattern vectors of low dimension.

The proposed method provides a closed form solution of the estimated expected value and estimated covariance matrix of the parameter vector. This parameter-estimation method can be implemented in a sensor network for real-time execution on limited-memory small micro-processors.

The major contributions of this paper are delineated below:

- Information fusion of observed evidence to obtain a statistical estimate of simultaneously varying parameters.
- Demonstration of possible non-uniqueness in parameter estimation, which may result due to selection of a scalar deviation measure as a cost functional in the multiple-parameter estimation problem.
- Robustness to process noise, sensor noise, and small fluctuations in parameter values, as discussed extensively in [18].
- Closed form solutions of estimated statistical parameters allowing for real-time execution on limited-memory platforms.

While there are many other issues that need to be addressed before the proposed estimation method can be considered for industrial applications, the following research topics are being currently pursued:

- Extension of the proposed method for estimation of a larger number (i.e., more than two) parameters in a variety of nonlinear dynamical systems.
- Extension of the parameter estimation problem under different types of nonlinearities with structured and unstructured uncertainties.
- Usage of the information on the state transition matrix for improved robustness of the parameter identification framework.

References

[1] Y. Ding, Z. Wu, Y. Zhang, Multi-fault diagnosis method based on a joint estimation of states and fault parameters, Journal of Tsinghua University 41 (12) (December 2001) 92–94.
[2] R. Isermann, Model-based fault-detection and diagnosis—status and applications, Annual Reviews in Control 29 (1) (2005) 71–85.
[3] A. Aitouche, D. Maquin, F. Busson, Multiple sensor fault detection in heat exchanger systems, Proceedings of the 1998 IEEE International Conference on Control Applications 2 (1–4 September 1998) 741–745.
[4] A. Ghosh, V. Kumar, B. Kulkarni, Parameter estimation in spatially extended systems: the Karhunen–Loeve and Galerkin multiple shooting approach, Physical Review E 64 (2001) 056222.
[5] H.-P. Huang, C.-C. Li, J.-C. Jeng, Multiple multiplicative fault diagnosis for dynamic processes via parameter similarity measures, Industrial & Engineering Chemistry Research 46 (13) (2007) 4517–4530.
[6] C. Koh, J. Shi, W. Williams, J. Ni, Multiple fault detection and isolation using the haar transform, part 1: theory, Journal of Manufacturing Science and Engineering 121 (2) (1999) 290–294.
[7] S. Gupta, A. Ray, S. Sarkar, M. Yasar, Fault detection and isolation in aircraft gas turbine engines: part I—the underlying concept, Proceedings of the IMechE, Part G—Journal of Aerospace Engineering 222 (3) (2008) 307–318.
[8] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, IEEE Proceedings 77 (2) (1989) 257–286.
[9] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Processing 84 (7) (2004) 1115–1130.
[10] 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).
[11] C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data, Review of Scientific Instruments 74 (2) (2003) 915–930.
[12] C. Rao, A. Ray, S. Sarkar, M. Yasar, Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns, 2008, doi: 10.1007/s11760-008-0061-8.
[13] X.Z. Tang, E.R. Tracy, R. Brown, Symbol statistics and spatio-temporal systems, Physica D: Nonlinear Phenomena 102 (3–4) (April 1997) 253–261.
[14] C. Piccardi, On parameter estimation of chaotic systems via symbolic time-series analysis, Chaos: An Interdisciplinary Journal of Nonlinear Science 16 (4) (2006) 043115.
[15] S. Mallat, A Wavelet Tour of Signal Processing, second ed., Academic Press, Boston, MA, 1998.
[16] L. Cohen, Time–Frequency Analysis, Prentice-Hall PTR, 1995.
[17] A. Subbu, A. Ray, Space partitioning via Hilbert transform for symbolic time series analysis, Applied Physics Letters 92 (8) (February 2008) 084107.
[18] V. Rajagopalan, A. Ray, Symbolic time series analysis via wavelet-based partitioning, Signal Processing 86 (11) (2006) 3309–3320.
[19] 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.
[20] V. Rajagopalan, S. Chakraborty, A. Ray, Estimation of slowly varying parameters in nonlinear systems via symbolic dynamic filtering, Signal Processing 88 (2) (February 2008) 339–348.
[21] J. Thompson, H. Stewart, Nonlinear Dynamics and Chaos, Wiley, Chichester, UK, 1986.
[22] M. Vidyasagar, Nonlinear Systems Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[23] D. Lind, M. Marcus, An Introduction to Symbolic Dynamics and Coding, Cambridge University Press, Cambridge, 1995.
[24] R.B. Bapat, T.E.S. Raghavan, Nonnegative Matrices and Applications, Cambridge University Press, Cambridge, 1997.
[25] H. Brunk, An Introduction to Mathematical Statistics, third ed., Xerox Publishing, Lexington, MA, 1995.
[26] R. Pathria, Statistical Mechanics, Elsevier Science and Technology Books, 1996.