Applied Soft Computing 37 (2015) 444–455

Contents lists available at ScienceDirect

Applied Soft Computing

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

Recursive subspace system identification for parametric fault detection in nonlinear systems

P. Gil[a,c,*], F. Santos[b], L. Palma[a], A. Cardoso[c]

[a] CTS-UNINOVA, Departamento de Engenharia Electrotécnica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

[b] Visteon Corporation Ltd, Electronics Product Group, 2951-503 Palmela, Portugal

[c] CISUC – Centre for Informatics and Systems of the University of Coimbra, Pólo II da Universidade de Coimbra, 3030-290 Coimbra, Portugal

ARTICLE INFO

Article history:
Received 8 October 2014
Received in revised form 30 June 2015
Accepted 19 August 2015
Available online 31 August 2015

Keywords:
Model-based fault detection
Parametric fault
Subspace system identification
Recursive estimation
Feedforward neural network

ABSTRACT

This work addresses the problem of detecting parametric faults in nonlinear dynamic systems by extending an eigenstructure based technique to a nonlinear context. Two local state-space models are updated online based on a recursive subspace system identification technique. One of the models relies on input–output real-time data collected from the plant, while the other is updated using data generated by a neural network predictor, describing the nonlinear plant behaviour in fault-free conditions. Parametric faults symptoms are generated based on eigenvalues residuals associated with two linear state-space model approximators. The feasibility and effectiveness of the proposed framework are demonstrated through two case studies.

1. Introduction

The increasing complexity and integration of industrial processes, some of them highly critical, has made it imperative to provide supervision systems with dedicated tools that could isolate and accommodate malfunctions or, generically faults, whenever needed. Furthermore, fault detection and isolation is a fundamental prerequisite for implementing conditioning-based maintenance procedures, in which the regular and systematic inspection of systems parts are replaced by analysing particular signals, along with decision-making actions that are performed on the basis of extracted features from data, either in real-time or offline.

Fault detection and isolation (FDI), as a whole, consists in making binary decisions concerning a given malfunctioning hypothesis and to determine its nature and location (see e.g. [14,3]). In general, FDI techniques rely on hardware-based schemes or on analytical redundancy approaches, or even on a combination of both. The former methodology is essentially based on comparing identical readings, collected with additional hardware, while the analytical or software based approaches make use of mathematical models and dedicated estimation methods (see e.g. [13,45]), in order to detect fault events. As this approach, commonly, does not require any additional hardware, its implementation is more attractive and cost-efficient. Nevertheless, analytical FDI methods are undoubtedly more challenging, as they need to cope with model uncertainties, unknown/unmeasurable disturbances and outliers, which may globally bias fault symptoms, and thus compromising the underlying sensitivity and specificity.

Among analytical approaches (see e.g. [21]), model-based fault detection and isolation techniques commonly resort to a set of residuals between a plant's readings and the outputs of a given predictor. By taking into account the residuals' magnitude and, possibly, other features, a classifier triggers symptoms regarding the presence or absence of faults. The residuals generation can be implemented based on state and output observers (see e.g. [44,31], parity relations (see e.g. [42,10,29]), or on parameters estimation using system identification techniques (see e.g. [43,7]). Concerning system identification based techniques, a model of the plant under normal operating conditions, assuming no faults, is derived either online or offline, and by detecting relevant changes in the model parameters (see e.g. [35]), the presence of faults are then isolated.

In a number of FDI problems the presence of parametric faults can be detected from changes in the eigenstructure of a linear state-space model describing the system dynamics. This model can be obtained from input–output data collected from the system

* Corresponding author at: CTS-UNINOVA, Departamento de Engenharia Electrotécnica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal. Tel.: +351 212 948 545; fax: +351 212 948 532.

E-mail addresses: psg@fct.unl.pt (P. Gil), fsantos9@visteon.com (F. Santos), lbp@fct.unl.pt (L. Palma), alberto@dei.uc.pt (A. Cardoso).

http://dx.doi.org/10.1016/j.asoc.2015.08.036
1568-4946/

and using, for instance, subspace-based linear system identification techniques (see e.g. [41]). In this case, as the derived model is described in state-space form, the eigenvalues associated with a given parameterization are immediately retrieved from the system matrix. Although this framework has been successfully applied in a number of case studies (see e.g. [2,1,8]), when the linearity assumption does not hold, the approach is doomed to fail, owing to the unreliability of residuals generation, which is partly related to model-plant mismatch. Hence, resorting to eigenvalues-based algorithms for FDI in a nonlinear context requires a completely different problem conceptualization and formulation.

The motivation and main contribution of this work is to extend a linear fault detection methodology relying on the system eigenstructure to nonlinear systems, for which faults are modelled as changes in the internal system dynamics. The approach makes use of a recursive subspace based system identification technique, along with the approximation capabilities of Nonlinear Autoregressive with Exogenous Inputs (NARX) neural networks, while assuming the input–output certainty equivalence principle. In this framework, two linear models are recursively updated in parallel. One of the models resorts to input–output data collected from the plant, while the other relies on input–output data provided by a NARX predictor. The corresponding eigenvalues are used to generate a set of residuals, from which symptoms of possible faults are triggered by a decision system.

The remainder of this paper is organized as follows. In Section 2 subspace-based methods for state-space system identification are discussed and two formulations presented, namely one where the parameters are estimated offline and the other based on a recursive implementation. Further, the model-based approach to fault detection is also discussed in terms of residuals and symptoms generation. Section 3 is devoted to presenting and describing the proposed approach for parametric fault detection in nonlinear systems, while Section 4 discusses some results obtained from two case studies. Finally, concluding remarks are drawn in Section 5.

2. System identification for fault detection

System identification deals with the problem of deriving an empirical model for a dynamical system based on input–output data. In the context of FDI, a model of the plant in normal or nominal operating conditions is first obtained. When a fault occurs, the underlying system behaviour in terms of outputs, inputs or internal dynamics, will differ from that predicted by the nominal model. As such, any fault event will lead to a change in the parameterization values.

The system identification problem aims at finding a relationship $g(\cdot)$ between past instantiations (u^{k-1}, y^{k-1}) and current outputs $y(k)$ (1), by appealing to a given regression technique, and taking into account an ordered data set sampled from the plant.

$$y(k) = g(u^{k-1}, y^{k-1}) + \vartheta(k) \tag{1}$$

where ϑ is an additive noise term and

$$\begin{aligned} \{u^{k-1}\} &\triangleq \begin{bmatrix} u(k-1) & \cdots & u(k-\alpha) \end{bmatrix}^{\mathrm{T}} \\ \{y^{k-1}\} &\triangleq \begin{bmatrix} y(k-1) & \cdots & y(k-\beta) \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{2}$$

with $\alpha, \beta \in \mathbb{N}^{+}$.

Among possible model structures $g(\cdot)$ to approximate the input–output behaviour of a plant, the present work considers the linear state-space models and a NARX neural network, with the choice of these structures dictated by the nature of the proposed FDI framework. The linear state-space model-based online identification relies on a recursive subspace technique, whereas the NARX neural network predictor training is carried out offline, using an iterative optimization algorithm.

2.1. Subspace system identification

A general feature of all Subspace System Identification (SID) methods is that they do not require *a priori* model parameterization, namely the model order, as its estimation is internally performed by the algorithm. Furthermore, the estimate of the underlying matrices relies on algebraic techniques, which makes them a very robust approach and less time consuming, compared to other methodologies, such as Prediction Error methods. Nevertheless, these methods can only provide suboptimal solutions (see e.g. [9,19]), which may ultimately impact on the approximation order and prediction performance.

Assume the linear time-invariant system described in state-space form as follows:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + \omega(k) \\ y(k) &= Cx(k) + Du(k) + \upsilon(k) \end{aligned} \tag{3}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^l$, $u \in \mathbb{R}^m$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{l\times n}$ and $D \in \mathbb{R}^{l\times m}$, while $\upsilon \in \mathbb{R}^l$ and $\omega \in \mathbb{R}^n$ are unobserved Gaussian distributed, zero mean, white noise sequences, accounting for the measurement noise and process noise, with covariances defined according to:

$$\mathbf{E}\left[\begin{pmatrix} \omega(p) \\ \upsilon(p) \end{pmatrix} \begin{pmatrix} \omega^T(q) & \upsilon^T(q) \end{pmatrix}\right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \tag{4}$$

with $\mathbf{E}$ denoting the expected value operator and δ_{pq} the Kronecker index. Moreover, suppose the available data collected from the plant are ergodic, the number of samples is sufficiently large $(N \to \infty)$, and Eq. (3) satisfies the following orthogonality property:

$$\mathbf{E}\left[\begin{pmatrix} x(k) \\ u(k) \end{pmatrix} \begin{pmatrix} \omega^T(k) & \upsilon^T(k) \end{pmatrix}\right] = 0 \tag{5}$$

2.1.1. Offline identification

Consider a data-set comprising an ordered sequence of input–output data collected from a plant, namely,

$$\begin{aligned} U^N &= \{u(0), u(1), \ldots, u(N-1)\} \\ Y^N &= \{y(1), y(2), \ldots, y(N)\} \end{aligned} \tag{6}$$

To come up with estimates for the state-space matrices (A, B, C, D) (up to within a similarity transformation) and error covariance matrices (Q, R, S), the data set $Z^N = \{U^{\mathrm{N}}, Y^{\mathrm{N}}\}$ is organized under the form of past and the future block Hankel matrices. For the input sequence U^{N}, the underlying block Hankel matrices take the following form:

$$U_p = \begin{pmatrix} u(0) & u(1) & \cdots & u(j-1) \\ u(1) & u(2) & \cdots & u(j) \\ \vdots & \vdots & \ddots & \vdots \\ u(i-1) & u(i) & \cdots & u(i+j-2) \end{pmatrix} \tag{7}$$

$$U_f = \begin{pmatrix} u(i) & u(i+1) & \cdots & u(i+j-1) \\ u(i+1) & u(i+2) & \cdots & u(i+j) \\ \vdots & \vdots & \ddots & \vdots \\ u(2i-1) & u(2i) & \cdots & u(2i+j-2) \end{pmatrix} \tag{8}$$