Journal of Process Control 38 (2016) 42–53

Contents lists available at ScienceDirect

Journal of Process Control

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

Integration of fault diagnosis and control based on a trade-off between fault detectability and closed loop performance

Yuncheng Du [a], Hector Budman [a], Thomas A. Duever [a,b,*]

[a] Department of Chemical Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1
[b] Department of Chemical Engineering, Ryerson University, Toronto, ON, Canada M5B 2K3

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form 1 September 2015
Accepted 8 December 2015
Available online 8 January 2016

Keywords:
Fault diagnosis
Bi-level optimization
Uncertainty analysis
Polynomial chaos
Control performance

ABSTRACT

This paper presents a novel methodology for simultaneous optimal tuning of a fault detection and diagnosis (FDD) algorithm and a feedback controller for a chemical plant in the presence of stochastic parametric faults. The key idea is to propagate the effect of time invariant stochastic uncertainties onto the measured variables by using a Generalized Polynomial Chaos (gPC) expansion and the nonlinear first principles' model of the process. A bi-level optimization is proposed for achieving a trade-off between the fault detectability and the closed loop process variability. The goal of the *outer level optimization* is to seek a trade-off between the efficiency of detecting a fault and the closed loop performance, while the *inner level optimization* is designed to optimally calibrate the FDD algorithm. The proposed method is illustrated by a continuous stirred tank reactor (CSTR) system with a fault consisting of stochastic and intermittent variations in the inlet concentration. Beyond achieving improved trade-offs between fault detectability and control, it is shown that the computational cost of the gPC model based method is lower than the Monte Carlo type sampling based approaches, thus demonstrating the potential of the gPC method for dealing with large problems and real-time applications.

1. Introduction

Equipment failures and abnormalities defined as faults are a major source of economic loss and safety hazards in many industries thus creating a need for fault detection and diagnosis algorithms. Most fault detection and diagnosis (FDD) systems are implemented at a supervisory hierarchical level above the control systems' level and use measured variables that are also used for feedback control. While there is a large body of literatures on FDD [1–6], the issue of integration between control and fault diagnosis algorithms has not been addressed as much in particular in the presence of stochastic faults.

A key challenge for integrating control and FDD is that they often have competing objectives. For instance, if the controlled variables are to be used for detection, better control means that the controlled variable deviates little from the set point, while FDD requires sufficiently large deviations for effective detection purposes [7,8]. Similar trade-offs occur also when the manipulated variables are used since good detection generally translate into large control actions as shown in this work. Moreover, process disturbances, nonlinearity and model error make the integration of FDD with control a challenging task [9]. Several methods have been proposed for optimal simultaneous tuning of a FDD algorithm and a controller based on robust norms. To synthesize the controller and diagnosis algorithms, a four parameter controller setup as a generalization of the two degrees of freedom controllers was proposed [10,11]. This method, however, did not explicitly address the cost of unobservable faults and their stochastic nature.

To improve the fault detectability in the presence of bounded uncertainties, set-based (separating inputs) FDD techniques have been used for active fault diagnosis [9,12,13]. These methods inject auxiliary signals into the system to enhance the detectability of faults. Instead of introducing an auxiliary signal in the current study, the controller is synthesized together with the fault detection algorithm.

Following the above, the current work addresses the problem of optimal simultaneous tuning of a FDD algorithm and the controller's parameters in the presence of time varying stochastic intermittent parametric faults, where the FDD is based on a nonlinear first principle model. The proposed approach seeks a trade-off between the fault detectability and the closed loop performance. Since the stochastic parametric faults (inputs) are considered, it is necessary to quantify the effect of these inputs on both the

* Corresponding author at: Department of Chemical Engineering, Ryerson University, Toronto, ON, Canada M5B 2K3. Tel.: +1 416 979 5140; fax: +1 416 979 5233.
E-mail address: tom.duever@ryerson.ca (T.A. Duever).

http://dx.doi.org/10.1016/j.jprocont.2015.12.007
0959-1524/

variables used in feedback control and for fault detection. One option to do such propagation and quantification is by Monte Carlo (MC) type sampling based simulations, which are computationally demanding since they require a large number of simulations of the nonlinear process model to get accurate results. Computational efficiency is critical in the current problem, since the propagation of the stochastic faults on other variables of interest has to be performed repetitively within the optimization algorithms used to achieve a trade-off between detection and control. Uncertainty analysis and propagation using the Generalized Polynomial Chaos (gPC) expansion has been studied by a number of authors in different areas, and has been reported to be more efficient as compared to MC simulations [14–18]. The advantage of gPC is that it can propagate a complex probability distribution into a variable of interest and explicitly calculate the statistics of the resulting outputs by analytical formulae [19,20].

The current work investigates the problem of optimal simultaneous tuning of a FDD algorithm and a controller in the presence of stochastic time varying disturbances by using the gPC expansions for stochastic parametric faults (inputs) and measured output variables. A bi-level optimization algorithm proposed in this work balances the fault detectability and the closed loop control performance. In both the works by Mesbah et al. [21] and our previous work [22] presented at the same meeting, the PDF profiles generated with the gPC models were utilized to enhance the fault detectability by minimizing the overlap between the PDF profiles. Unlike the referenced work [21], the previous study done by the authors [22] and the current work synthesize the fault detection algorithm together with the controller to seek an optimal trade-off between detection and control. Also, the current work differs from previous studies in the proposed fault detection algorithm that it is based on a maximum likelihood criterion to detect the fault using a gPC model. Preliminary results of seeking a trade-off between the fault detectability and the closed loop control performance were outlined in [22]. A significant reduction in computational effort was observed by using the gPC method, as compared with the MC sampling based approaches, which is further investigated in this work. Also, the earlier work by the authors [22] is extended by combining the gPC theory with the maximum likelihood based estimation to recursively estimate the stochastic parametric faults (inputs) during transients, while in our previous study only the steady state fault detection problem was considered. The application of the gPC model with maximum likelihood dynamically estimates the value of the stochastic fault over a time moving window. The estimation results can be used as a real-time process monitoring strategy for detection of stochastic faults in nonlinear systems. While previously reported parameter estimation approaches based on combinations of the gPC with Bayesian and maximum likelihood have been applied in an offline fashion [23–26], the current work proposes a gPC based methodology for online detection of faults.

To summarize, the novel contributions in this current work are: (i) The use, in the context of integration between fault diagnosis and control, of an intrusive gPC approach for uncertainty propagation and quantification by substituting the gPC directly into the first principles nonlinear model of the system; (ii) The use of the maximum likelihood based estimation in combination with the gPC model for fault detection; and (iii) The formulation of a bi-level optimization for achieving an optimal tradeoff between control and improved fault detection. The methodology is specifically targeted to: (i) Balance the control performance and the fault detectability, by synthesizing a FDD algorithm that is operated together with a feedback controller; and (ii) Diagnose the stochastic faults consisting of uncertainties around mean values that change intermittently, using measurements collected immediately after the occurrence of a step change on the mean values of the faults.

This paper is organized as follows. Section 2 presents the background and the principal methodologies used in this work. The optimization problems formulated for simultaneously tuning the FDD algorithm and the controller are given in Section 3. The presentation of the maximum likelihood based FDD algorithm is also presented in Section 3. An endothermic continuous stirred tank reactor (CSTR) is introduced as a case study in Section 4. Analysis and discussion of the results are presented in Section 5 followed by conclusions in Section 6.

2. Generalized Polynomial Chaos expansion

The Generalized Polynomial Chaos (gPC) expansion [20] represents an arbitrary continuous random variable of interest as a polynomial series of another random variable with a given standard distribution. Assume a set of nonlinear ordinary differential equations (ODEs) describe the dynamic behavior of a system:

$$\begin{aligned} &\dot{\boldsymbol{x}} = f(t, \boldsymbol{x}, \boldsymbol{u}; \boldsymbol{g}) \\ &0 \le t \le t_f, \quad \boldsymbol{x}(0) \le \boldsymbol{x}_0 \end{aligned}, \tag{1}$$

where the vector $\boldsymbol{x} \in \boldsymbol{R}^n$ contains the system states (measured variables) with initial conditions $\boldsymbol{x}_0 \in \boldsymbol{R}^n$ over time domain $[0, t_f]$, and $\boldsymbol{u}$ denotes the known inputs of the system. The vector $\boldsymbol{g} \in \boldsymbol{R}^{n_g}$ is the unknown stochastic time varying input. Note that this work assumes that the input vector $\boldsymbol{g}$ is the stochastic parametric faults of interest. The '·' notation over $\boldsymbol{x}$ signifies the derivative with respect to time t. The function f is assumed to be the first principle model of the process. To quantify the effect of stochastic inputs (faults) $\boldsymbol{g}$ on the different measured variables, the gPC expansion can be employed. To that purpose each unknown input g_i ($i = 1,2,\ldots,n_g$) in g is represented as a function of a set of random variables $\boldsymbol{\xi} = \{\boldsymbol{\xi}_i\}$:

$$g_i = g_i(\boldsymbol{\xi}_i), \tag{2}$$

where $\boldsymbol{\xi}_i$ is the ith random variable. The random variables ($\boldsymbol{\xi} = \{\boldsymbol{\xi}_i\}$) are assumed to be independent and identically distributed. Following the gPC expansion, the unknown stochastic faults (inputs) $\boldsymbol{g}(\boldsymbol{\xi})$ and system states $\boldsymbol{x}(t,\boldsymbol{\xi})$ are described in terms of orthogonal polynomial basis functions $\Phi_k(\boldsymbol{\xi})$:

$$g(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} g_k \Phi_k(\boldsymbol{\xi}) \tag{3}$$

$$\boldsymbol{x}(t, \boldsymbol{\xi}) = \sum_{k=0}^{\infty} x_k(t) \Phi_k(\boldsymbol{\xi}), \tag{4}$$

where x_k and g_k are the gPC coefficients of measured variables (states) and faults at each time instant t, $\Phi_k(\boldsymbol{\xi})$ are multi-dimensional orthogonal basis functions of $\boldsymbol{\xi}$ in the gPC theory. If the input ($\boldsymbol{g}$) can be measured or estimated, the coefficients of the unknown input, g_k, can be calculated such that (3) follows an a priori measured statistical distribution. Then, the gPCs representing the measured quantities (states) responses resulting from this random input can be calculated using a model of the process combined with a Galerkin projection procedure [19]. By Galerkin projection it is possible to compute the expansion coefficients $\{x_k(t)\}$ by projecting (1) onto each one of the polynomial chaos basis functions$\{\Phi_k(\boldsymbol{\xi})\}$ as described in (5):

$$\langle \dot{\boldsymbol{x}}(t, \boldsymbol{\xi}), \Phi_k(\boldsymbol{\xi}) \rangle = \langle f(t, \boldsymbol{x})(t, \boldsymbol{\xi}), \boldsymbol{u}(t), \boldsymbol{g}(\boldsymbol{\xi}), \Phi_k(\boldsymbol{\xi}) \rangle \tag{5}$$

For practical application, (3) and (4) are often truncated to a finite number of terms, i.e., P. Hence, the total number of terms in (5) is a function of an arbitrary order p in (3) that is necessary to represent an a priori known distribution of $\boldsymbol{g}$ and the number (n_g) of different faults (inputs) in vector $\boldsymbol{g}$ as follows:

$$P = ((n_g + p)!/(n_g!p!)) - 1 \tag{6}$$