Automatica 74 (2016) 107–117

Contents lists available at ScienceDirect

Automatica

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

Closed-loop input design for guaranteed fault diagnosis using set-valued observers☆

Davide M. Raimondo[a], Giuseppe Roberto Marseglia[a], Richard D. Braatz[b], Joseph K. Scott[c]

[a] Identification and Control of Dynamic Systems Laboratory, University of Pavia, Italy
[b] Massachusetts Institute of Technology, Cambridge, MA, USA
[c] Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, SC, USA

ARTICLE INFO

Article history:
Received 14 December 2015
Accepted 17 June 2016

Keywords:
Fault detection
Set-based estimation
Zonotopes
Robust control

ABSTRACT

Active fault diagnosis (AFD) can be used to improve the diagnosability of faults by injecting a suitably designed input into a process. When faults are described as discrete switches between linear systems with uncertainties bounded within zonotopes, an optimal open-loop input guaranteeing diagnosis within a specified time horizon can be computed efficiently by solving a Mixed Integer Quadratic Program (MIQP). In this article, the constrained zonotope (CZ) set representation recently developed by the authors is used to extend the MIQP approach to general polytopic uncertainties without sacrificing efficiency. Next, this approach is combined with a CZ-based set-valued observer in a moving horizon framework to achieve rigorous closed-loop AFD. This method can greatly accelerate diagnosis relative to the open-loop approach, but requires online optimization. To reduce the online cost, we propose a method for solving the open-loop problem explicitly with respect to past measurements and inputs, which requires only observability of the nominal and faulty models. The effectiveness of the proposed approaches is demonstrated through several numerical examples.

1. Introduction

The detection and diagnosis of malfunctions and other abnormal events (i.e., faults) is an essential control task for engineered systems in the chemical, power, aerospace, and mechanical domains (Gao, Cecati, & Ding, 2015; Tchakoua et al., 2014; Yu, Woradechjumroen, & Yu, 2014). Without corrective action, faults can lead to performance degradation and potentially critical situations. However, fault detection and diagnosis are challenging due to the presence of disturbances, measurement noises, and the actions of feedback controllers. Approaches to automatic fault diagnosis can be classified as either active or passive. In the passive approach, input–output data are collected in real-time and faults are diagnosed based on comparisons with a process model or historical data. In contrast, the active approach involves injecting a signal into the system to improve the diagnosability of potential faults with minimal impact on the nominal system (Gao et al., 2015).

This article considers input design for active fault diagnosis of linear systems subject to bounded process and measurement noise, and faults modeled by discrete changes in the system matrices. Several works have addressed this problem using inputs that excite specially designed residual signals in the presence of faults (Kerestecioglu & Cetin, 2004; Niemann, 2006). A multi-model stochastic formulation is considered in Blackmore, Rajamanoharan, and Williams (2008) and Cheong and Manchester (2015), where inputs are designed to minimize the probability of incorrect diagnosis. A similar approach for nonlinear systems is given in Streif, Petzke, Mesbah, Findeisen, and Braatz (2014). Several multi-model formulations with deterministic bounds on the measurement and process noises have also been proposed. Interestingly, these either provide an input that is guaranteed to identify the correct model within a specified time horizon, or conclude that no such input exists. The article (Nikoukhah & Campbell, 2006) considers noises that are energy-bounded within ellipsoids. Pointwise-in-time polytopic bounds are considered in Nikoukhah (1998), but costly computations with high-dimensional

☆ The material in this paper was presented at the 12th European Control Conference, July 17–19, 2013, Zurich, Switzerland. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

E-mail addresses: davide.raimondo@unipv.it (D.M. Raimondo), groberto.marseglia@gmail.com (G. Roberto Marseglia), braatz@mit.edu (R.D. Braatz), jks9@clemson.edu (J.K. Scott).

http://dx.doi.org/10.1016/j.automatica.2016.07.033
0005-1098/

polytopes are required. More recently, an efficient method using pointwise bounds described by zonotopes was developed, making it tractable solve diagnosis problems with high dimension and/or multiple fault models (Scott, Findeisen, Braatz, & Raimondo, 2014). Extensions of the preceding approaches include methods for nonlinear systems (Andjelkovic, Sweetingham, & Campbell, 2008; Paulson, Raimondo, Braatz, Findeisen, & Streif, 2014), hybrid stochastic–deterministic approaches (Marseglia, Scott, Magni, Braatz, & Raimondo, 2014; Scott, Marseglia, Magni, Braatz, & Raimondo, 2013), methods with input and robust state constraints (Andjelkovic & Campbell, 2011; Scott et al., 2014), and robust MPC with diagnosis constraints (Raimondo, Marseglia, Braatz, & Scott, 2013).

The above approaches are all open-loop in the sense that the computed active input is applied with no online modification. The design of exogenous active inputs for closed-loop systems has been studied (Ashari, Nikoukhah, & Campbell, 2012a,b). However, the feedback law was given *a priori*, not designed for fault diagnosis. In Stoican, Olaru, Seron, and De Doná (2012), the authors propose a fault tolerant control method which relies on the computation of invariant sets and a reference governor scheme to isolate faults. While the feedback law was also given a priori, the reference was suitably chosen to guarantee the separation of the residual sets for the healthy and faulty dynamics. In Niemann, Stoustrup, and Poulsen (2014), a given feedback controller is temporarily modified to make a residual more sensitive to faults. A closed-loop approach for stochastic models is presented in Puncochar, Siroky, and Simandl (2015) and Simandl and Puncochar (2009), where the input minimizes nominal control objectives and risks associated with incorrect diagnosis. Finally, a deterministic closed-loop approach using polytopes is described in Tabatabaeipour (2015).

In this context, the present article makes three main contributions. First, the open-loop input design method in Scott et al. (2014) is generalized. This method provides guaranteed diagnosis for linear multi-model systems with initial conditions, disturbances, and measurement noises bounded pointwise by zonotopes. Here, the *constrained zonotope* computations recently proposed in Scott, Marseglia, Raimondo, and Braatz (2016) are used to extend this approach to general polytopic uncertainties, while maintaining the efficiency of the original approach (see Section 3). Second, a new closed-loop input design method is developed by applying the open-loop method of Section 3 within a moving horizon framework, where online measurements are incorporated through set-valued observers (Section 4). This method potentially provides much less conservative active inputs on average (e.g., reduced length, norm), while maintaining the guarantee of fault diagnosis within a given time horizon. Among existing closed-loop approaches, such a guarantee is only provided by the method in Tabatabaeipour (2015). However, that method uses polytope projection operations that scale exponentially in the system dimension (Althoff, Stursberg, & Buss, 2010; Fukuda, 2004). Numerical experiments in Scott, Findeisen, Braatz, and Raimondo (2013) clearly show that such projections are intractable for systems with more than 2 or 3 states. In contrast, our use of constrained zonotopes here avoids this computation completely. Our third contribution is a method for computing an explicit feedback law off-line for cases where computing open-loop inputs online is prohibitive (Section 5). This is enabled by the use of *finite-memory* set-valued observers, at the cost of some additional conservatism. Compared to our preliminary results in Raimondo, Braatz, and Scott (2013), the closed-loop approaches here use more effective observers (enabled by the developments of Section 3), and the explicit method is generalized to address the case of incomplete state measurements.

1.1. Problem formulation

Consider a discrete-time system with time k, state $\mathbf{x}_k \in \mathbb{R}^{n_x}$, output $\mathbf{y}_k \in \mathbb{R}^{n_y}$, input $\mathbf{u}_k \in \mathbb{R}^{n_u}$, disturbance $\mathbf{w}_k \in \mathbb{R}^{n_w}$, and measurement error $\mathbf{v}_k \in \mathbb{R}^{n_v}$. In each interval $[k, k+1]$, $k = 0, 1, \ldots$, the system evolves according to one of n_m possible linear models. The matrices of these models are distinguished by the argument $i \in \mathbb{I} \equiv \{1, \ldots, n_m\}$:

$$\mathbf{x}_{k+1} = \mathbf{A}(i)\mathbf{x}_k + \mathbf{B}(i)\mathbf{u}_k + \mathbf{r}(i) + \mathbf{B}_w(i)\mathbf{w}_k, \quad (1)$$

$$\mathbf{y}_k = \mathbf{C}(i)\mathbf{x}_k + \mathbf{s}(i) + \mathbf{D}_v(i)\mathbf{v}_k. \quad (2)$$

The model $i = 1$ is nominal, and the rest are faulty. Models representing multiple, simultaneous faults can be included in $\mathbb{I}$ if desired (Scott et al., 2014). The constant vectors $\mathbf{r}(i)$ and $\mathbf{s}(i)$ are used to model additive faults such as sensor and actuator bias. Let $\mathbf{x}_0 \in X_0(i)$ represent any information known about $\mathbf{x}_0$ prior to $k = 0$, given that model $i \in \mathbb{I}$ is active. $X_0(i)$ can depend on i if, e.g., it has been constructed from previous measurements through (2). We assume that $(\mathbf{w}_k, \mathbf{v}_k) \in W \times V$, $\forall k \in \mathbb{N}$, and that W, V, and $X_0(i)$ are bounded convex polytopes. Our objective is to design input sequences that guarantee fault diagnosis over a finite horizon N. Specifically, assuming that one model $i^* \in \mathbb{I}$ is active on $[0, N]$ (i.e., the test interval), we aim to design an input that can identify i^* with certainty, while simultaneously satisfying convex polytopic constraints $\mathbf{u}_k \in U$, $\forall k \in \mathbb{N}$, and minimizing a quadratic cost function. The proposed methods are appropriate for designing short test signals that are applied periodically, or after a fault has been detected but not diagnosed.

2. Preliminaries

2.1. Constrained zonotopes and set operations

The new methods in this article are largely enabled by computations with *constrained zonotopes*, a new class of sets introduced in Scott et al. (2016) as an extension of the zonotopes.

Definition 1. A set $Z \subset \mathbb{R}^n$ is a *constrained zonotope* if there exists $(\mathbf{G}, \mathbf{c}, \mathbf{A}, \mathbf{b}) \in \mathbb{R}^{n\times n_g} \times \mathbb{R}^n \times \mathbb{R}^{n_c\times n_g} \times \mathbb{R}^{n_c}$ such that

$$Z = \{\mathbf{G}\boldsymbol{\xi} + \mathbf{c} : \|\boldsymbol{\xi}\|_\infty \leq 1,\ \mathbf{A}\boldsymbol{\xi} = \mathbf{b}\}. \quad (3)$$

In contrast to standard zonotopes, Definition 1 permits linear equality constraints on $\boldsymbol{\xi}$. The columns of $\mathbf{G}$ are called the *generators*, $\mathbf{c}$ is the *center*, and $\mathbf{A}\boldsymbol{\xi} = \mathbf{b}$ are the *constraints*. We use the shorthand $Z = \{\mathbf{G}, \mathbf{c}, \mathbf{A}, \mathbf{b}\}$ and $Z = \{\mathbf{G}, \mathbf{c}\}$ for constrained and standard zonotopes, respectively.

Constrained zonotopes are substantially more flexible than zonotopes. Indeed, a central result in Scott et al. (2016) is that Z is a constrained zonotope iff it is a convex polytope; i.e., iff Z is bounded and $\exists(\mathbf{H}, \mathbf{k}) \in \mathbb{R}^{n_h\times n} \times \mathbb{R}^n$ such that Z can be written in the halfspace representation (H-rep) $Z = \{\mathbf{z} \in \mathbb{R}^n : \mathbf{Hz} \leq \mathbf{k}\}$. We refer to (3) as the *constrained generator representation* (CG-rep) of Z. Converting from H- to CG-rep is simple, and while the converse is difficult in general, it is never required in the proposed methods (Scott et al., 2016).

The CG-rep has two primary advantages compared to the H-rep. First, it trivializes the computation of some important set operations. Let $Z, W \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^k$, $\mathbf{R} \in \mathbb{R}^{k\times n}$, and define

$$\mathbf{R}Z \equiv \{\mathbf{Rz} : \mathbf{z} \in Z\}, \quad (4)$$

$$Z \oplus W \equiv \{\mathbf{z} + \mathbf{w} : \mathbf{z} \in Z,\ \mathbf{w} \in W\}, \quad (5)$$

$$Z \cap_{\mathbf{R}} Y \equiv \{\mathbf{z} \in Z : \mathbf{Rz} \in Y\}. \quad (6)$$

Eq. (4) is a linear mapping of Z, (5) is the *Minkowski sum*, and (6) is a generalized intersection that arises in state estimation (see