Automatica 50 (2014) 1580–1589

Contents lists available at ScienceDirect

Automatica

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

Input design for guaranteed fault diagnosis using zonotopes☆

Joseph K. Scott [a,1], Rolf Findeisen [b], Richard D. Braatz [a], Davide M. Raimondo [c]

[a] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[b] Institute for Automation Engineering, Laboratory for Systems Theory and Automatic Control, Otto-von-Guericke University Magdeburg, Germany

[c] Identification and Control of Dynamic Systems Laboratory, University of Pavia, Italy

ARTICLE INFO

Article history:
Received 26 February 2013
Received in revised form
7 February 2014
Accepted 26 February 2014
Available online 24 April 2014

Keywords:
Fault diagnosis
Input design
Reachability analysis
Zonotopes

ABSTRACT

An input design method is presented for guaranteeing the diagnosability of faults from the outputs of a system. Faults are modeled by discrete switches between linear models with bounded disturbances and measurement errors. Zonotopes are used to efficiently characterize the set of inputs that are guaranteed to lead to outputs that are consistent with at most one fault scenario. Provided that this set is nonempty, an element is then chosen that is minimally harmful with respect to other control objectives. This approach leads to a nonconvex optimization problem, but is shown to be equivalent to a mixed-integer quadratic program that can be solved efficiently. Methods are given for reducing the complexity of this program, including an observer-based method that drastically reduces the number of binary variables when many sampling times are required for diagnosis.

1. Introduction

In many industries (chemical (Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003), aerospace (Zolghadri, 2010), etc.), the trend toward increasing complexity and automation has made component malfunctions and other abnormal events (i.e., faults) increasingly frequent. At the same time, economic considerations have led to the use of inexpensive and unreliable components in many mass market applications. Accordingly, achieving safe and reliable operation for many systems now requires fast and accurate methods for detecting and diagnosing faults on the basis of process measurements. These tasks are rendered difficult by the confounding effects of disturbances, measurement uncertainty, and the compensatory actions of the control system.

Fault detection and diagnosis methods can be categorized as either passive or active. Passive approaches attempt to diagnose faults by comparing the available input–output data for the process to models or historical data (Chiang, Russell, & Braatz, 2001; Venkatasubramanian et al., 2003). Often, however, faults may not be detectable in the available measurements, or cannot be diagnosed without exciting the system. Accordingly, the active approach involves injecting a signal into the system to improve detectability of the fault (Blackmore, Rajamanoharan, & Williams, 2008; Campbell & Nikoukhah, 2004; Esna Ashari, Nikoukhah, & Campbell, 2012; Niemann, 2006; Nikoukhah, 1998; Simandl & Puncochar, 2009).

This article presents a set-based approach for active fault diagnosis. The process of interest, under nominal and various faulty conditions, is described by a set of linear discrete-time models subject to bounded disturbances and measurement errors. Faults are modeled by discrete switches between these models. The proposed framework permits multiple faults occurring either sequentially or simultaneously, although computational complexity ultimately limits the number of scenarios considered (see Section 2). Given a set of scenarios, the objective is to compute an input that is guaranteed to generate outputs consistent with at most one scenario, thereby providing a complete fault diagnosis. Such inputs are referred to as *separating inputs*. In addition to this diagnosis condition, the computed input is further required to be minimally harmful with respect to other control objectives.

This problem was first considered in Nikoukhah (1998). In the case of two models (one nominal and one faulty), the set of separating inputs was shown to be the complement of a projection of a high-dimensional polytope. Unfortunately, polytope projection is computationally intensive and numerically unstable in the required dimensions. The book (Campbell & Nikoukhah,

☆ The material in this paper was partially presented at the 2013 American Control Conference (ACC 2013), June 17–19, 2013, Washington, DC, USA. This paper was recommended for publication in revised form by Associate Editor Michele Basseville under the direction of Editor Torsten Söderström. BP is acknowledged for financial support of this project.

E-mail addresses: jkscott@mit.edu (J.K. Scott), rolf.findeisen@ovgu.de (R. Findeisen), braatz@mit.edu (R.D. Braatz), davide.raimondo@unipv.it (D.M. Raimondo).

[1] Tel.: +1 8646560997; fax: +1 8646560784.

http://dx.doi.org/10.1016/j.automatica.2014.03.016
0005-1098/

2004) proposes an active input design method for the case where the disturbances and measurement errors are energy bounded rather than pointwise bounded. The input is chosen as the solution of a bilevel optimization problem in which the outer program searches for a minimum two-norm input and the inner program restricts the feasible set to separating inputs. This optimization problem is nonconvex and is solved in the two-model case by a specialized algorithm. Various extensions of this approach have been investigated, including methods for continuous-time and nonlinear systems (Andjelkovic, Sweetingham, & Campbell, 2008; Campbell & Nikoukhah, 2004), asymptotically optimal implementations (Nikoukhah & Campbell, 2006; Nikoukhah, Campbell, & Delebecque, 2000), and methods for systems under linear feedback control (Ashari, Nikoukhah, & Campbell, 2009, 2012; Esna Ashari et al., 2012). A more general optimization formulation has also been proposed that permits multiple fault models and arbitrary objectives and constraints (Campbell, Horton, & Nikoukhah, 2002; Campbell & Nikoukhah, 2004). However, the structure of the two-model formulation is lost and the method instead relies on general-purpose software to solve difficult optimal control problems constrained by two-point boundary-value problems.

This article treats the case where the disturbances and measurement errors are pointwise bounded rather than energy bounded, and uses zonotopes rather than polytopes or ellipsoids. After a formal problem statement and some preliminary developments in Sections 2 and 3, the set of separating inputs is characterized using efficient zonotope computations in Section 4, effectively eliminating the polytope projection problem in Nikoukhah (1998). This result is then used to pose a bilevel optimization problem for choosing an optimal separating input in Section 5, similar to the approach in Nikoukhah and Campbell (2006). The use of zonotopes here permits a reformulation as a mixed-integer quadratic program (MIQP) for which the number of integer variables can be controlled using zonotope order reduction techniques. The resulting optimization problem is simple to implement and practically solvable, while being flexible with respect to the choice of objective, the presence of state and control constraints, the possibility of multiple fault models, and the possibility of multiple faults occurring simultaneously or sequentially in the time interval of interest. Techniques for reducing the computational complexity of the approach are discussed in Section 6, and an approximate implementation of the approach using set-valued observers is proposed in Section 7 to reduce the complexity when many sampling times are required for diagnosis. Numerical examples are presented in Section 8, and Section 9 contains concluding remarks. This article extends the preliminary results in Scott, Findeisen, Braatz, and Raimondo (2013) by providing a more general theoretical development, an improved optimization formulation, a treatment of state constraints, several new methods for reducing computational complexity, and extended numerical results.

2. 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 $i \in \mathbb{I} \equiv \{1, \ldots, n_m\}$:

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

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

The model $i = 1$ is nominal, and the rest are faulty. Models representing multiple, simultaneous faults can be included in $\mathbb{I}$ if desired. The constant vectors $\mathbf{r}(i)$ and $\mathbf{s}(i)$ are used to model additive faults such as sensor and actuator bias. It is assumed that $\mathbf{x}_0 \in X_0$, and $(\mathbf{w}_k, \mathbf{v}_k) \in W \times V$, $\forall k \in \mathbb{N}$, where X_0, W and V are zonotopes (see Section 3.1).

A fault at time k is modeled by a transition from one model in $\mathbb{I}$ to another; i.e., $i_k \neq i_{k-1}$. Given a time interval $[0, N]$, a *fault scenario* on $[0, N]$ is defined as a sequence $(i_0, \ldots, i_N) \in \mathbb{I}^N$. Let $\tilde{\mathbb{I}} \subset \mathbb{I}^N$ denote a set of permissible fault scenarios on $[0, N]$. Given $\tilde{\mathbb{I}}$, the goal is to compute an open-loop input sequence $\tilde{\mathbf{u}} = (\mathbf{u}_0, \ldots, \mathbf{u}_{N-1})$ such that any observed sequence of outputs $\tilde{\mathbf{y}} = (\mathbf{y}_0, \ldots, \mathbf{y}_N)$ is consistent with at most one fault scenario in $\tilde{\mathbb{I}}$, regardless of the exact values of the initial condition, disturbances, and measurement errors in the sets X_0, W, and V. Such input sequences are referred to as *separating inputs* (see Section 4). Ideally, $\tilde{\mathbf{u}}$ should be *minimal* in some sense (e.g., length, norm). We assume that N is specified and focus on the computation of a separating input sequence that minimizes a quadratic objective subject to input and state constraints. This computation can be iterated with N increasing from 1 until the problem becomes feasible.

Requiring that $\tilde{\mathbf{u}}$ is a separating input is equivalent to requiring that every distinct pair of scenarios $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ can be distinguished. Thus, $\tilde{\mathbf{u}}$ must satisfy $Q = \binom{s}{2}$ conditions, where s is the number of scenarios in $\tilde{\mathbb{I}}$ (see (19) in Section 5). Despite the computational advantages of the proposed methods, this combinatorial dependence demands a parsimonious selection of permissible scenarios. If every scenario is permissible, then $s = (n_m)^N$. However, many scenarios will be nonsensical (e.g., spontaneously corrected faults) or very unlikely (e.g., multiple unrelated faults). Further reductions can be achieved by limiting the frequency of faults (i.e., imposing a minimum number of repeats $i_k = i_{k+1} = \cdots = i_{k+d}$ after a transition). Effectively choosing scenarios for a given application is not considered here; $\tilde{\mathbb{I}}$ is assumed given.

3. Preliminaries

3.1. Zonotopes and set operations

The methods in this article involve computations with *zonotopes*, which are centrally symmetric convex polytopes that can be described as Minkowski sums of line segments (Kuhn, 1998). In *generator representation*, a zonotope Z is prescribed by its *center* $\mathbf{c} \in \mathbb{R}^n$ and *generators* $\mathbf{g}_1, \ldots, \mathbf{g}_{n_g} \in \mathbb{R}^n$ as $Z = \{\mathbf{G}\boldsymbol{\xi} + \mathbf{c} : \boldsymbol{\xi} \in \mathbb{R}^{n_g},\ \|\boldsymbol{\xi}\|_\infty \leq 1\}$, where $\mathbf{G} \equiv [\mathbf{g}_1 \ \ldots \ \mathbf{g}_{n_g}]$. We use the notation $Z = \{\mathbf{G}, \mathbf{c}\}$. The *order* of a zonotope is n_g/n. A first-order zonotope with linearly independent generators is a parallelotope.

Let $Z, Y \subset \mathbb{R}^n$, $\mathbf{R} \in \mathbb{R}^{m \times n}$, and define the operations

$$\mathbf{R}Z \equiv \{\mathbf{R}\mathbf{z} : \mathbf{z} \in Z\}, \tag{3}$$

$$Z \oplus Y \equiv \{\mathbf{z} + \mathbf{y} : \mathbf{z} \in Z,\ \mathbf{y} \in Y\}, \tag{4}$$

$$Z \ominus Y \equiv \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} + Y \subset Z\}. \tag{5}$$

When $Z = \{\mathbf{G}_z, \mathbf{c}_z\}$ and $Y = \{\mathbf{G}_y, \mathbf{c}_y\}$ are zonotopes, (3)–(4) are also zonotopes and can be computed efficiently (Kuhn, 1998):

$$\mathbf{R}Z = \{\mathbf{R}\mathbf{G}_z, \mathbf{R}\mathbf{c}_z\}, \qquad Z \oplus Y = \{[\mathbf{G}_z\ \mathbf{G}_y], \mathbf{c}_z + \mathbf{c}_y\}. \tag{6}$$

In contrast, when Z and Y are general convex polytopes, the Minkowski sum (4) and the linear mapping (3) with singular $\mathbf{R}$ (e.g., polytope projection) both become extremely computationally demanding and numerically unstable in dimensions greater than about 10 (Althoff & Krogh, 2011; Fukuda, 2004). However, the results of the operations in (6) can be higher order than Z and Y. To avoid increasing orders, techniques for enclosing a given zonotope within a zonotope of lower order must be used. For the computations presented in Section 8, Method C in Althoff, Stursberg, and Buss (2010) is used.

When Y is a zonotope and Z is a polytope in the standard half-space representation, the Pontryagin difference (5) can be computed easily as follows (see Theorem 2.3 in Kolmanovsky & Gilbert, 1998).

Lemma 1. *Let* $Y = \{\mathbf{G}, \mathbf{c}\}$ *and* $Z \equiv \{\mathbf{z} \in \mathbb{R}^n : \mathbf{H}\mathbf{z} \leq \mathbf{k}\}$, *where* $\mathbf{k} \in \mathbb{R}^m$ *and* $\mathbf{H} \in \mathbb{R}^{m\times n}$ *has rows* $\mathbf{h}_i^{\mathrm{T}} \neq \mathbf{0}$. *Then* $Z \ominus Y = \{\mathbf{z} \in \mathbb{R}^n : \mathbf{H}\mathbf{z} \leq \mathbf{k}'\}$, *where* $k_i' \equiv k_i - \mathbf{h}_i^{\mathrm{T}}\mathbf{c} - \|\mathbf{h}_i^{\mathrm{T}}\mathbf{G}\|_1,\ i = 1, \ldots, m$.

Note that every zonotope satisfies $\{\mathbf{G}, \mathbf{c}\} = \{-\mathbf{G}, \mathbf{c}\}$ and $\{\mathbf{G}, \mathbf{c}\} = \mathbf{c} \oplus \{\mathbf{G}, \mathbf{0}\}$. Moreover, the following result holds (see Lemma 2.1 in Dobkin, Hershberger, Kirkpatrick, & Suri, 1993).

Lemma 2. *Let* $Z = \{\mathbf{G}_z, \mathbf{a}_z + \mathbf{b}_z\}$ *and* $Y = \{\mathbf{G}_y, \mathbf{a}_y + \mathbf{b}_y\}$. *Then* $Z \cap Y = \emptyset$ *if and only if* $\mathbf{a}_y - \mathbf{a}_z \notin \{\mathbf{G}_z, \mathbf{b}_z\} \oplus \{\mathbf{G}_y, -\mathbf{b}_y\}$.

3.2. Reachable set notation and computations

Reachable sets for (1)–(2) will be used to characterize the set of separating inputs in Section 4. Below, a tilde designates a sequence associated with (1)–(2), e.g., $\tilde{\mathbf{u}} = (\mathbf{u}_0, \ldots, \mathbf{u}_{N-1}) \in \mathbb{R}^{Nn_u}$. For $0 \leq \ell \leq k < N$, we further denote $\tilde{\mathbf{u}}_{\ell:k} = (\mathbf{u}_\ell, \ldots, \mathbf{u}_k)$. Similarly, $\tilde{i} = (i_0, \ldots, i_N) \in \tilde{\mathbb{I}}$ denotes an arbitrary fault scenario.

For $k \in \mathbb{N}$, define the solution mappings

$$(\phi_k, \psi_k) : \mathbb{R}^{kn_u} \times \mathbb{I}^{k+1} \times \mathbb{R}^{n_x} \times \mathbb{R}^{kn_w} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$$

so that $\phi_k(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{w}}, \mathbf{v})$ and $\psi_k(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{w}}, \mathbf{v})$ are the state and output of (1)–(2) at k, respectively, given the specified inputs. Strictly, ϕ_k does not depend on i_k and $\mathbf{v}$ (see (1)–(2)), but they are included for notational convenience. Let $\tilde{\phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) = (\phi_\ell, \ldots, \phi_k)$ and $\tilde{\psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{\mathbf{w}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{v}}) = (\psi_\ell, \ldots, \psi_k)$, where we have abbreviated $(\phi_j, \psi_j) = (\phi_j, \psi_j)(\tilde{\mathbf{u}}_{0:j-1}, \tilde{i}_{0:j}, \mathbf{x}_0, \tilde{\mathbf{w}}_{0:j-1}, \mathbf{v}_j)$, $\ell \leq j \leq k$.

For each $\tilde{i} \in \mathbb{I}^k$ and $\tilde{\mathbf{u}} \in \mathbb{R}^{kn_u}$, define the *reachable state and output sets* on $[\ell, k]$ by

$$\tilde{\Phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) \equiv \{\tilde{\phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) : (\mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) \in X_0 \times \tilde{W} \times \tilde{V}\},$$
$$\tilde{\Psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) \equiv \{\tilde{\psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) : (\mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) \in X_0 \times \tilde{W} \times \tilde{V}\},$$

where $\tilde{W} = W \times \cdots \times W$ and $\tilde{V} = V \times \cdots \times V$ with $k-\ell$ and $k-\ell+1$ Cartesian products, respectively. Explicit dependence on X_0, $\tilde{W}$, and $\tilde{V}$ is omitted for brevity. The reachable state and output sets *at* k are defined as $\Phi_k(\tilde{\mathbf{u}}, \tilde{i}) \equiv \tilde{\Phi}_{k:k}(\tilde{\mathbf{u}}, \tilde{i})$ and $\Psi_k(\tilde{\mathbf{u}}, \tilde{i}) \equiv \tilde{\Psi}_{k:k}(\tilde{\mathbf{u}}, \tilde{i})$.

Note that (1)–(2) recursively define matrices $\tilde{\mathbf{A}}(\tilde{i}), \tilde{\mathbf{B}}(\tilde{i})$, etc., which depend on ℓ and k, such that

$$\tilde{\phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) = \tilde{\mathbf{A}}(\tilde{i})\mathbf{x}_0 + \tilde{\mathbf{B}}(\tilde{i})\tilde{\mathbf{u}} + \tilde{\mathbf{r}}(\tilde{i}) + \tilde{\mathbf{B}}_w(\tilde{i})\tilde{\mathbf{w}},$$
$$\tilde{\psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) = \tilde{\mathbf{C}}(\tilde{i})\tilde{\phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) + \tilde{\mathbf{s}}(\tilde{i}) + \tilde{\mathbf{D}}_v(\tilde{i})\tilde{\mathbf{v}}.$$

Using (3)–(4), it follows that

$$\tilde{\Phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) = \tilde{\mathbf{A}}(\tilde{i})X_0 \oplus \tilde{\mathbf{B}}(\tilde{i})\tilde{\mathbf{u}} \oplus \tilde{\mathbf{r}}(\tilde{i}) \oplus \tilde{\mathbf{B}}_w(\tilde{i})\tilde{W}, \tag{7}$$
$$\tilde{\Psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) = \tilde{\mathbf{C}}(\tilde{i})\tilde{\Phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) \oplus \tilde{\mathbf{s}}(\tilde{i}) \oplus \tilde{\mathbf{D}}_v(\tilde{i})\tilde{V}. \tag{8}$$

Denote $X_0 = \{\mathbf{G}_0, \mathbf{c}_0\}$, $W = \{\mathbf{G}_W, \mathbf{c}_W\}$, and $V = \{\mathbf{G}_V, \mathbf{c}_V\}$. It is simple to see that $\tilde{W}$ is a zonotope with center $\mathbf{c}_{\tilde{W}} = (\mathbf{c}_W, \ldots, \mathbf{c}_W)$ and block-diagonal generator matrix $\mathbf{G}_{\tilde{W}} = \mathrm{diag}(\mathbf{G}_W, \ldots, \mathbf{G}_W)$, and $\tilde{V}$ is analogous. Thus, the above reachable sets can be computed efficiently using (6). Moreover, these sets take the form

$$\tilde{\Phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) = \{\mathbf{G}^{\Phi}_{\ell:k}(\tilde{i}), \overline{\phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i})\}, \tag{9}$$
$$\tilde{\Psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) = \{\mathbf{G}^{\Psi}_{\ell:k}(\tilde{i}), \overline{\psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i})\}, \tag{10}$$

with the generator matrices $\mathbf{G}^{\Phi}_{\ell:k}(\tilde{i}) = [\tilde{\mathbf{A}}(\tilde{i})\mathbf{G}_0\ \tilde{\mathbf{B}}_w(\tilde{i})\mathbf{G}_{\tilde{W}}]$ and $\mathbf{G}^{\Psi}_{\ell:k}(\tilde{i}) = [\tilde{\mathbf{C}}(\tilde{i})\mathbf{G}^{\Phi}_{\ell:k}(\tilde{i})\ \tilde{\mathbf{D}}_v(\tilde{i})\mathbf{G}_{\tilde{V}}]$, and the centers $\overline{\phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) = \tilde{\phi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{c}_0, \mathbf{c}_{\tilde{W}}, \mathbf{c}_{\tilde{V}})$ and $\overline{\psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}) = \tilde{\psi}_{\ell:k}(\tilde{\mathbf{u}}, \tilde{i}, \mathbf{c}_0, \mathbf{c}_{\tilde{W}}, \mathbf{c}_{\tilde{V}})$.

When $\ell = k = N$ in (7)–(8), the sets $\Phi_N(\tilde{\mathbf{u}}, \tilde{i})$ and $\Psi_N(\tilde{\mathbf{u}}, \tilde{i})$ can also be computed recursively as the state and output at N, respectively, of the iteration

$$X_{k+1} = \mathbf{A}(i_k)X_k \oplus \mathbf{B}(i_k)\mathbf{u}_k \oplus \mathbf{r}(i_k) \oplus \mathbf{B}_w(i_k)W, \tag{11}$$
$$Y_k = \mathbf{C}(i_k)X_k \oplus \mathbf{s}(i_k) \oplus \mathbf{D}_v(i_k)V. \tag{12}$$

Moreover, the explicit dependence of the results on $\tilde{\mathbf{u}}$ can be computed by noting that $\overline{\phi}_N(\tilde{\mathbf{u}}, \tilde{i}) = \overline{\phi}_N(\mathbf{0}, \tilde{i}) + \overline{\mathbf{B}}^{\Phi}_N(\tilde{i})\tilde{\mathbf{u}}$ and $\overline{\psi}_N(\tilde{\mathbf{u}}, \tilde{i}) = \overline{\psi}_N(\mathbf{0}, \tilde{i}) + \overline{\mathbf{B}}^{\Psi}_N(\tilde{i})\tilde{\mathbf{u}}$, where $\overline{\mathbf{B}}^{\Phi}_N(\tilde{i})$ and $\overline{\mathbf{B}}^{\Psi}_N(\tilde{i})$ are the N^{th} elements of the recursion

$$\overline{\mathbf{B}}^{\Phi}_{k+1} = [\mathbf{A}(i_k)\overline{\mathbf{B}}^{\Phi}_k\ \mathbf{B}(i_k)], \qquad \overline{\mathbf{B}}^{\Psi}_k = \mathbf{C}(i_k)\overline{\mathbf{B}}^{\Phi}_k. \tag{13}$$

In contrast, the sets $\tilde{\Phi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i})$ and $\tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i})$, given by choosing $\ell = 0$ and $k = N$ in (7)–(8), cannot be computed recursively. Instead, the set operations in (7)–(8) must be carried out exactly as written, which requires that the high-dimensional matrices $\tilde{\mathbf{A}}(\tilde{i}), \tilde{\mathbf{B}}(\tilde{i})$, etc., are explicitly constructed and stored. These calculations may be prohibitive when Nn_x is very large (see Section 7). On the other hand, the calculations maintain the coupling between all states (respectively, outputs) through the initial condition and past disturbances, i.e.,

$$\tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i}) \subset \Psi_0(i_0) \times \cdots \times \Psi_N(\tilde{\mathbf{u}}, \tilde{i}). \tag{14}$$

4. Separating inputs

Suppose the input $\tilde{\mathbf{u}} = (\mathbf{u}_0, \ldots, \mathbf{u}_{N-1})$ is injected and the output $\tilde{\mathbf{y}} = (\mathbf{y}_0, \ldots, \mathbf{y}_N)$ is observed. From the preceding section, it is clear that $\tilde{\mathbf{y}}$ is consistent with scenario $\tilde{i} \in \tilde{\mathbb{I}}$ iff

$$\tilde{\mathbf{y}} \in \tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i}). \tag{15}$$

Thus, we are interested in inputs that ensure that (15) holds for at most one $\tilde{i} \in \tilde{\mathbb{I}}$. For any such input, checking (15) for each $\tilde{i} \in \tilde{\mathbb{I}}$ online either provides the desired fault diagnosis (including the nominal case), or concludes that no scenario in $\tilde{\mathbb{I}}$ is correct.

Definition 1. An input $\tilde{\mathbf{u}} \in \mathbb{R}^{Nn_u}$ *separates* $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ *on* $[0, N]$ if

$$\tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i}) \cap \tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{j}) = \emptyset. \tag{16}$$

Similarly, $\tilde{\mathbf{u}}$ *separates* $\tilde{\mathbb{I}}$ *on* $[0, N]$, or is a *separating input*, if it separates every $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ with $\tilde{i} \neq \tilde{j}$. The set of all separating inputs is denoted by $\mathscr{S}(\tilde{\mathbb{I}})$.

If (16) fails, we may still observe $\tilde{\mathbf{y}} \in \tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i}) \setminus \tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{j})$. Thus, an input that is not separating can lead to a diagnosis if particular values of the outputs are observed. Similarly, a separating input on $[0, N]$ can potentially return a guaranteed diagnosis after only $M < N$ steps (Campbell, Drake, & Nikoukhah, 2002).

In the interest of satisfying (16), a critical observation is that $\tilde{\mathbf{u}}$ only translates $\tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{j})$ with respect to $\tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i})$, and does not change the shape or relative orientation of these sets, which follows directly from (10) and leads to a simple characterization of $\mathscr{S}(\tilde{\mathbb{I}})$:

Theorem 3. *An input* $\tilde{\mathbf{u}}$ *belongs to* $\mathscr{S}(\tilde{\mathbb{I}})$ *if and only if*

$$\tilde{\mathbf{N}}(\tilde{i}, \tilde{j})\tilde{\mathbf{u}} \notin \tilde{Z}(\tilde{i}, \tilde{j}), \tag{17}$$

for all $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}},\ \tilde{i} \neq \tilde{j}$, *where* $\tilde{\mathbf{N}}(\tilde{i}, \tilde{j}) \equiv (\tilde{\mathbf{C}}(\tilde{j})\tilde{\mathbf{B}}(\tilde{j}) - \tilde{\mathbf{C}}(\tilde{i})\tilde{\mathbf{B}}(\tilde{i}))$ *and* $\tilde{Z}(\tilde{i}, \tilde{j}) \equiv \{[\mathbf{G}^{\Psi}_{0:N}(\tilde{i})\ \mathbf{G}^{\Psi}_{0:N}(\tilde{j})], \overline{\psi}_{0:N}(\mathbf{0}, \tilde{i}) - \overline{\psi}_{0:N}(\mathbf{0}, \tilde{j})\}$.

Proof. Apply Lemma 2 to (16) using (10) and note that $\overline{\psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i}) = \tilde{\mathbf{C}}(\tilde{i})\tilde{\mathbf{B}}(\tilde{i})\tilde{\mathbf{u}} + \overline{\psi}_{0:N}(\mathbf{0}, \tilde{i})$.

Eq. (17) and the compactness of each $\tilde{Z}(\tilde{i},\tilde{j})$ implies that the set of inputs separating $\tilde{i},\tilde{j}\in\tilde{\mathbb{I}}$ is nonempty and unbounded whenever $\tilde{\mathbf{N}}(\tilde{i},\tilde{j})\neq\mathbf{0}$. If $\tilde{\mathbf{N}}(\tilde{i},\tilde{j})=\mathbf{0}$, then $\tilde{i}$ and $\tilde{j}$ are separated by every input if $\mathbf{0}\notin\tilde{Z}(\tilde{i},\tilde{j})$, and are not separated by any input otherwise.

When X_0, W, and V are general convex polytopes, the results in Nikoukhah (1998) show that the set of inputs separating $\tilde{i}$ and $\tilde{j}$ is the complement of a convex polytope P. When X_0, W, and V are zonotopes, Theorem 3 shows that these inputs can be alternatively characterized by the requirement that $\tilde{\mathbf{N}}(\tilde{i},\tilde{j})\tilde{\mathbf{u}}$ lies in the complement of a zonotope. In the context of optimization over $\mathscr{S}(\tilde{\mathbb{I}})$, the latter characterization has several advantages. Most importantly, the data required by (17) (i.e. $\tilde{\mathbf{N}}(\tilde{i},\tilde{j})$ and $\tilde{Z}(\tilde{i},\tilde{j})$) can be efficiently computed using the developments of Section 3.2. In contrast, the polytope P in Nikoukhah (1998) is an Nn_u-dimensional projection of a $N(n_y+n_u)$-dimensional polytope, and its computation becomes intractable for $N(n_y+n_u)$ greater than about 10. Additionally, the set of inputs separating $\tilde{i}$ and $\tilde{j}$ is nonconvex, since both characterizations ultimately require $\tilde{\mathbf{u}}$ to lie in the complement of a convex set. However, it is shown in Section 5 that (17) can be reformulated as a mixed-integer linear constraint in which the number of binary variables can be favorably controlled using the zonotopic structure.

5. Optimal separating inputs

This section considers the problem of selecting a particular element of $\mathscr{S}(\tilde{\mathbb{I}})$ by minimizing a given quadratic objective function, subject to given input constraints. To this end, let $\mathbf{R}\in\mathbb{R}^{n_u\times n_u}$ be positive semidefinite, let $U\subset\mathbb{R}^{n_u}$ be a convex polytope, and define $J(\tilde{\mathbf{u}})\equiv\sum_{k=0}^{N-1}(\mathbf{u}_k)^{\mathrm{T}}\mathbf{R}\mathbf{u}_k$ and $\tilde{U}\equiv U\times\cdots\times U$. An optimal separating input is chosen by solving

$$\inf\{J(\tilde{\mathbf{u}}):\tilde{\mathbf{u}}\in\tilde{U}\cap\mathscr{S}(\tilde{\mathbb{I}})\}. \tag{18}$$

The purpose of this optimization is to minimize any harmful effects of the separating input with respect to other control objectives. One simple formulation is to find the minimum two-norm separating input. More generally, using the separation condition as a constraint in a predictive control formulation also results in an optimization of the form (18).

The optimization (18) is difficult owing to nonconvexity of $\mathscr{S}(\tilde{\mathbb{I}})$. To arrive at an efficient solution method, we invoke Theorem 3 and carry out a series of reformulations leading to a mixed-integer quadratic program (MIQP). Let $q\in\{1,\ldots,Q\}$ index the combinations $\tilde{i},\tilde{j}\in\tilde{\mathbb{I}}$ with $\tilde{i}\neq\tilde{j}$, and denote $\tilde{\mathbf{N}}^{[q]}\equiv\tilde{\mathbf{N}}(\tilde{i},\tilde{j})$ and $\tilde{Z}^{[q]}\equiv\tilde{Z}(\tilde{i},\tilde{j})$. According to Theorem 3, (18) is equivalent to

$$\inf\{J(\tilde{\mathbf{u}}):\tilde{\mathbf{u}}\in\tilde{U},\ \tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}}\notin\tilde{Z}^{[q]},\ q=1,\ldots,Q\}. \tag{19}$$

The following lemma shows that the separation constraints in (19) can be formulated as linear programs. Denote $\tilde{Z}^{[q]}=\{\mathbf{G}^{[q]},\mathbf{c}^{[q]}\}$, where $\mathbf{G}^{[q]}$ has $n_g^{[q]}$ columns. It is assumed for simplicity that each $\mathbf{G}^{[q]}$ has full row rank, which implies that each $\tilde{Z}^{[q]}$ has a nonempty interior and holds if, for example, $\mathbf{D}_v(i)\mathbf{G}_V$ has full row rank for all $i\in\mathbb{I}$ (see Remark 1).

Lemma 4. *For each $\tilde{\mathbf{u}}\in\tilde{U}$ and $q\in\{1,\ldots,Q\}$, define*

$$\widehat{\delta}^{[q]}(\tilde{\mathbf{u}})\equiv\min_{\delta^{[q]},\boldsymbol{\xi}^{[q]}}\ \delta^{[q]} \tag{20}$$

$$s.t.\ \tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}}=\mathbf{G}^{[q]}\boldsymbol{\xi}^{[q]}+\mathbf{c}^{[q]},\qquad\|\boldsymbol{\xi}^{[q]}\|_\infty\leq 1+\delta^{[q]}.$$

Then $\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}}\notin\tilde{Z}^{[q]}$ iff $\widehat{\delta}^{[q]}(\tilde{\mathbf{u}})>0$.

Proof. Since $\mathbf{G}^{[q]}$ is full row rank, (20) is feasible. If $\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}}\notin\tilde{Z}^{[q]}$, then $\nexists\boldsymbol{\xi}$ such that $\|\boldsymbol{\xi}\|_\infty\leq 1$ and $\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}}=\mathbf{G}^{[q]}\boldsymbol{\xi}^{[q]}+\mathbf{c}^{[q]}$. Thus, (20) has no feasible point with $\delta^{[q]}\leq 0$. Conversely, if $\widehat{\delta}^{[q]}(\tilde{\mathbf{u}})>0$, then there cannot exist a feasible point of (20) with $\delta^{[q]}\leq 0$, and hence $\nexists\boldsymbol{\xi}$ such that $\|\boldsymbol{\xi}\|_\infty\leq 1$ and $\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}}=\mathbf{G}^{[q]}\boldsymbol{\xi}^{[q]}+\mathbf{c}^{[q]}$.

From the condition $\widehat{\delta}^{[q]}(\tilde{\mathbf{u}})>0$ in Lemma 4, it is evident that the feasible set in (19) is not closed, and there may not exist $\tilde{\mathbf{u}}^*$ attaining the infimum. For this reason, the constraints $\widehat{\delta}^{[q]}(\tilde{\mathbf{u}})\geq\varepsilon$ will be enforced instead, where $\varepsilon>0$ is a *minimum separation threshold* (see (21)). Although this introduces a small amount of conservatism, any $\tilde{\mathbf{u}}$ satisfying these constraints is still a separating input. For the remaining reformulations, it is assumed that an upper bound $\widehat{\delta}_m^{[q]}$ is available such that $\widehat{\delta}^{[q]}(\tilde{\mathbf{u}})\leq\widehat{\delta}_m^{[q]}$, $\forall\tilde{\mathbf{u}}\in\tilde{U}$ (see Section 5.1). Then (19) can be written as the bilevel program

$$\min\{J(\tilde{\mathbf{u}}):\tilde{\mathbf{u}}\in\tilde{U},\ \varepsilon\leq\widehat{\delta}^{[q]}(\tilde{\mathbf{u}})\leq\widehat{\delta}_m^{[q]},\ q=1,\ldots,Q\}. \tag{21}$$

A single level program is obtained by replacing the linear inner programs (20) with their necessary and sufficient conditions of optimality (see Proposition 3.4.1 in Bertsekas, 1999):

$$\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}}=\mathbf{G}^{[q]}\boldsymbol{\xi}^{[q]}+\mathbf{c}^{[q]}, \tag{22}$$

$$\|\boldsymbol{\xi}^{[q]}\|_\infty\leq 1+\delta^{[q]}, \tag{23}$$

$$(\mathbf{G}^{[q]})^{\mathrm{T}}\boldsymbol{\lambda}^{[q]}=(\boldsymbol{\mu}_1^{[q]}-\boldsymbol{\mu}_2^{[q]}), \tag{24}$$

$$1=(\boldsymbol{\mu}_1^{[q]}+\boldsymbol{\mu}_2^{[q]})^{\mathrm{T}}\mathbf{1}, \tag{25}$$

$$\mathbf{0}\leq\boldsymbol{\mu}_1^{[q]},\boldsymbol{\mu}_2^{[q]}, \tag{26}$$

$$0=\mu_{1,k}^{[q]}(\xi_k^{[q]}-1-\delta^{[q]}),\quad\forall k=1,\ldots,n_g^{[q]}, \tag{27}$$

$$0=\mu_{2,k}^{[q]}(\xi_k^{[q]}+1+\delta^{[q]}),\quad\forall k=1,\ldots,n_g^{[q]}. \tag{28}$$

In (25), $\mathbf{1}$ is a vector of ones. For each q, the constraint $\varepsilon\leq\widehat{\delta}^{[q]}(\tilde{\mathbf{u}})\leq\widehat{\delta}_m^{[q]}$ in (21) can now be replaced by the condition: $\exists(\delta^{[q]},\boldsymbol{\xi}^{[q]},\boldsymbol{\lambda}^{[q]},\boldsymbol{\mu}_1^{[q]},\boldsymbol{\mu}_2^{[q]})$ such that $\varepsilon\leq\delta^{[q]}\leq\widehat{\delta}_m^{[q]}$ and (22)–(28) hold. However, the complementarity constraints (27)–(28) are nonconvex. Thus, we introduce binary variables $\mathbf{p}_1^{[q]},\mathbf{p}_2^{[q]}\in\{0,1\}^{n_g^{[q]}}$ and replace (27)–(28) with the implications (Fortuny-Amat & McCarl, 1981):

$$\begin{aligned}
&p_{1,k}^{[q]}=1\Longrightarrow\mu_{1,k}^{[q]}\text{free},\qquad(\xi_k^{[q]}-1-\delta^{[q]})=0,\\
&p_{1,k}^{[q]}=0\Longrightarrow\mu_{1,k}^{[q]}=0,\qquad(\xi_k^{[q]}-1-\delta^{[q]})\text{free},\\
&p_{2,k}^{[q]}=1\Longrightarrow\mu_{2,k}^{[q]}\text{free},\qquad(\xi_k^{[q]}+1+\delta^{[q]})=0,\\
&p_{2,k}^{[q]}=0\Longrightarrow\mu_{2,k}^{[q]}=0,\qquad(\xi_k^{[q]}+1+\delta^{[q]})\text{free}.
\end{aligned} \tag{29}$$

These implications can be enforced through the linear constraints:

$$\mu_{1,k}^{[q]}\leq p_{1,k}^{[q]},\qquad\mu_{2,k}^{[q]}\leq p_{2,k}^{[q]}, \tag{30}$$

$$(\xi_k^{[q]}-1-\delta^{[q]})\in[-2(1+\widehat{\delta}_m^{[q]})(1-p_{1,k}^{[q]}),0], \tag{31}$$

$$(\xi_k^{[q]}+1+\delta^{[q]})\in[0,2(1+\widehat{\delta}_m^{[q]})(1-p_{2,k}^{[q]})]. \tag{32}$$

In particular, the bounds imposed by (30) when $p_{1,k}^{[q]}=1$ or $p_{2,k}^{[q]}=1$ are consequences of (26) and (25), and hence are not restrictive. Similarly, the bounds imposed by (31) (respectively, (32)) when $p_{1,k}^{[q]}=0$ (resp. $p_{2,k}^{[q]}=0$) are consequences of (23) and $\delta^{[q]}\leq\delta_m^{[q]}$. Thus, (21) is equivalent to

$$\min_{\tilde{\mathbf{u}},\delta^{[q]},\boldsymbol{\xi}^{[q]},\boldsymbol{\lambda}^{[q]},\boldsymbol{\mu}_1^{[q]},\boldsymbol{\mu}_2^{[q]},\mathbf{p}_1^{[q]},\mathbf{p}_2^{[q]}}J(\tilde{\mathbf{u}}) \tag{33}$$

$$\text{s.t. }\tilde{\mathbf{u}}\in\tilde{U},$$

$$\left\{\begin{array}{l}\varepsilon\leq\delta^{[q]}\leq\delta_m^{[q]},\ (22)\text{–}(26),\\ \mathbf{p}_1^{[q]},\mathbf{p}_2^{[q]}\in\{0,1\}^{n_g^{[q]}},\ (30)\text{–}(32)\end{array}\right\}\quad\forall q\in\{1,\ldots,Q\}.$$

This is a MIQP and can be solved efficiently using, for example, CPLEX (IBM, 2012).

Remark 1. If $\mathbf{G}^{[q]}$ is not full row rank for some q, then the conclusion of Lemma 4 holds upon replacing the constraints in (20) by $\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}}=\mathbf{G}^{[q]}\boldsymbol{\xi}^{[q]}+\mathbf{H}^{[q]}\boldsymbol{\gamma}^{[q]}+\mathbf{c}^{[q]}$, $\|\boldsymbol{\xi}^{[q]}\|_\infty\leq 1+\delta^{[q]}$, and

$\|\boldsymbol{\gamma}^{[q]}\|_\infty \leq \delta^{[q]}$, where $\mathbf{H}^{[q]}$ is a matrix with minimal number of columns such that $[\mathbf{G}^{[q]}\ \mathbf{H}^{[q]}]$ has full row rank. A modified set of optimality conditions (22)–(28) can then be derived, and the subsequent reformulations can be repeated to yield an analog of (33).

Remark 2. In Nikoukhah and Campbell (2006), a method is given for computing separating inputs when $(\mathbf{x}_0, \tilde{\mathbf{w}}, \tilde{\mathbf{v}}) \in E$ for an ellipsoid E. However, the related problem with independent, pointwise ellipsoidal bounds $\mathbf{x}_0 \in X_0$, $\mathbf{w}_k \in W$, and $\mathbf{v}_k \in V$, $\forall k \in \mathbb{N}$, has not been addressed. Since zonotopes can outer-approximate ellipsoids with arbitrary precision, separating inputs for this case can be computed using the methods above, with some conservatism. Under these assumptions, the reachable sets $\tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i})$, and moreover the sets $\tilde{Z}^{[q]}$, are not ellipsoids. However, they are *zonoids* (i.e., limits of zonotopes) (Bolker, 1969), and can therefore be approximated as zonotopes.

5.1. Computing $\widehat{\delta}_m^{[q]}$

The above reformulations leading to (30)–(32) required a bound $\widehat{\delta}_m^{[q]}$ such that $\widehat{\delta}_m^{[q]} \geq \widehat{\delta}_{m,*}^{[q]} \equiv \max\{\widehat{\delta}^{[q]}(\tilde{\mathbf{u}}) : \tilde{\mathbf{u}} \in \tilde{U}\}$. Since $\widehat{\delta}^{[q]}$ is defined by (20) as the parametric solution of a LP, Theorem 5.1 in Bertsimas and Tsitsiklis (1997) ensures that it is convex on $\tilde{U}$. Thus, $\widehat{\delta}_{m,*}^{[q]}$ can be computed as the maximum value of $\widehat{\delta}^{[q]}$ over all vertices $\tilde{\mathbf{u}}_v$ of $\tilde{U}$. If the number of vertices is prohibitive, an alternative approach is to consider

$$\widehat{\delta}_{m,g+}^{[q]} \equiv \sup\{\widehat{\delta}^{[q]}(\tilde{\mathbf{u}}) : \tilde{\mathbf{u}} \in \tilde{U},\ \widehat{\delta}^{[q]}(\tilde{\mathbf{u}}) \leq \widehat{\delta}_{m,g}^{[q]}\}, \tag{34}$$

where $\widehat{\delta}_{m,g}^{[q]}$ is a guess for $\widehat{\delta}_{m,*}^{[q]}$. This is a bilevel program with (20) embedded, but is easier to solve than $\max\{\widehat{\delta}^{[q]}(\tilde{\mathbf{u}}) : \tilde{\mathbf{u}} \in \tilde{U}\}$ because $\widehat{\delta}_{m,g}^{[q]}$ provides the upper bound necessary to reformulate it as an MILP using the techniques of the previous section. Rather than providing $\widehat{\delta}_{m,*}^{[q]}$ directly, the solution of (34) determines if the guess $\widehat{\delta}_{m,g}^{[q]}$ is valid, and returns the optimal bound $\widehat{\delta}_{m,g+}^{[q]} = \widehat{\delta}_{m,*}^{[q]}$ whenever it is:

Lemma 5. *$\widehat{\delta}_{m,g+}^{[q]} < \widehat{\delta}_{m,g}^{[q]}$ iff $\widehat{\delta}_{m,*}^{[q]} < \widehat{\delta}_{m,g}^{[q]}$, and in this case $\widehat{\delta}_{m,g+}^{[q]} = \widehat{\delta}_{m,*}^{[q]}$.*

Proof. If $\widehat{\delta}_{m,*}^{[q]} < \widehat{\delta}_{m,g}^{[q]}$, then (34) is equivalent to $\max\{\widehat{\delta}^{[q]}(\tilde{\mathbf{u}}) : \tilde{\mathbf{u}} \in \tilde{U}\}$, and hence $\widehat{\delta}_{m,g+}^{[q]} = \widehat{\delta}_{m,*}^{[q]} < \widehat{\delta}_{m,g}^{[q]}$. Conversely, if $\widehat{\delta}_{m,*}^{[q]} \geq \widehat{\delta}_{m,g}^{[q]}$, then (34) is a true restriction of $\max\{\widehat{\delta}^{[q]}(\tilde{\mathbf{u}}) : \tilde{\mathbf{u}} \in \tilde{U}\}$, and it is clear that $\widehat{\delta}_{m,g+}^{[q]} = \widehat{\delta}_{m,g}^{[q]}$.

Although computing $\widehat{\delta}_{m,*}^{[q]}$ through (34) requires solving at least one MILP, it avoids explicit enumeration of the vertices of $\tilde{U}$. Also, solving (34) is much cheaper than solving (33) because each instance of (34) has only one embedded LP, and hence many fewer binary variables.

5.2. State constraints

In many cases, it may be desirable to restrict (18) by requiring that the system states remain robustly within a given polytope $\mathscr{X} \subset \mathbb{R}^{n_x}$, regardless of the correct scenario:

$$\Phi_k(\tilde{\mathbf{u}}_{0:k-1}, \tilde{i}_{0:k}) \subset \mathscr{X}, \quad \forall k \in \{0, \ldots, N\},\ \forall \tilde{i} \in \tilde{\mathbb{I}}. \tag{35}$$

The zonotopic structure of Φ_k allows this restriction to be easily expressed by a system of linear constraints on $\tilde{\mathbf{u}}$ as follows. For each k and $\tilde{i}$, (35) can be expressed using (5) and (9) as $\overline{\phi}_k(\tilde{\mathbf{u}}_{0:k-1}, \tilde{i}_{0:k}) \in \mathscr{X} \ominus \{\mathbf{G}_k^{\Phi}(\tilde{i}_{0:k}), \mathbf{0}\}$. Given $\mathscr{X} = \{\mathbf{z} : \mathbf{H}\mathbf{z} \leq \mathbf{k}\}$, Lemma 1 can be used to compute $\mathbf{k}'$ such that $\{\mathbf{z} : \mathbf{H}\mathbf{z} \leq \mathbf{k}'\} = \mathscr{X} \ominus \{\mathbf{G}_k^{\Phi}(\tilde{i}_{0:k}), \mathbf{0}\}$. Thus, (35) is equivalent to $\mathbf{H}\overline{\phi}_k(\tilde{\mathbf{u}}_{0:k-1}, \tilde{i}_{0:k}) \leq \mathbf{k}'$, which is a polyhedral constraint on $\tilde{\mathbf{u}}_{0:k-1}$ after recursively computing the nominal state $\overline{\phi}_k(\tilde{\mathbf{u}}_{0:k-1}, \tilde{i}_{0:k})$ as an affine function of $\tilde{\mathbf{u}}_{0:k-1}$.

6. Computational complexity and approximations

The complexity of (33) is governed by the number of binary variables, which is $B = \sum_{q=1}^{Q}(2n_g^{[q]})$. Assuming that $n_g^{[q]} = n_g$ for all q and letting $\ell = n_g/((N+1)n_y)$ be the order of the zonotopes $\tilde{Z}^{[q]}$, $B = Q(2(N+1)n_y\ell)$. This section and the next consider methods for reducing complexity by reducing either Q or $(2(N+1)n_y\ell)$. In all of these methods, the requirements on $\tilde{\mathbf{u}}$ are tightened. Thus, a more conservative input may result, but the guarantee of diagnosis is maintained.

6.1. Pair elimination

To each distinct pair of scenarios in $\tilde{\mathbb{I}}$, there corresponds a constraint $\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}} \notin \tilde{Z}^{[q]} \equiv \{\mathbf{G}^{[q]}, \mathbf{c}^{[q]}\}$, leading to $(2(N+1)n_y\ell)$ binary variables in (33). In this section, a preprocessing step is developed that, for each pair, attempts to eliminate this constraint or replace it by a single linear constraint on $\tilde{\mathbf{u}}$, thus eliminating $(2(N+1)n_y\ell)$ binary variables.

For each pair $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$, indexed by q, we solve the QP

$$\min\{J(\tilde{\mathbf{u}}) : \tilde{\mathbf{u}} \in \tilde{U},\ \tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}} = \mathbf{G}^{[q]}\boldsymbol{\xi} + \mathbf{c}^{[q]},\ \|\boldsymbol{\xi}\|_\infty \leq 1\}. \tag{36}$$

If (36) is infeasible, then $\nexists \tilde{\mathbf{u}} \in \tilde{U}$ such that $\mathbf{N}^{[q]}\tilde{\mathbf{u}} \in Z^{[q]}$, and hence the separation constraint for this pair can be eliminated from (19). Otherwise, let $\tilde{\mathbf{u}}_*^{[q]}$ and $J_*^{[q]}$ be optimal solution and objective values and consider the constraint

$$(\lambda^{[q]})^{\mathrm{T}}\tilde{\mathbf{u}} \leq (1-\gamma)J_*^{[q]}, \quad (\lambda^{[q]})^{\mathrm{T}} \equiv (\tilde{\mathbf{u}}_*^{[q]})^{\mathrm{T}}\tilde{\mathbf{R}}, \tag{37}$$

where $\gamma > 0$ is a specified threshold and $\tilde{\mathbf{R}}$ is the matrix such that $J(\tilde{\mathbf{u}}) = \tilde{\mathbf{u}}^{\mathrm{T}}\tilde{\mathbf{R}}\tilde{\mathbf{u}}$.

Lemma 6. *If $\tilde{\mathbf{u}} \in \tilde{U}$ does not separate $\tilde{i}$ and $\tilde{j}$ on $[0, N]$, then it violates (37). Moreover, every $\tilde{\mathbf{u}}$ violating (37) has $J(\tilde{\mathbf{u}}) \geq (1-\gamma)^2 J_*^{[q]}$.*

Proof. Choose $\tilde{\mathbf{u}}$ in the set $\{\tilde{\mathbf{u}} \in \tilde{U} : \mathbf{N}^{[q]}\tilde{\mathbf{u}} \in Z^{[q]}\}$. Since this set is convex, it contains all convex combinations of $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{u}}_*^{[q]}$. Thus, by optimality of $\tilde{\mathbf{u}}_*^{[q]}$,

$$(\tilde{\mathbf{u}}_*^{[q]} + \sigma(\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_*^{[q]}))^{\mathrm{T}}\tilde{\mathbf{R}}(\tilde{\mathbf{u}}_*^{[q]} + \sigma(\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_*^{[q]})) \geq J_*^{[q]}, \tag{38}$$

for all $\sigma \in [0, 1]$. Some rearrangement gives that

$$2\sigma(\lambda^{[q]})^{\mathrm{T}}(\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_*^{[q]}) + \sigma^2 J(\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_*^{[q]}) \geq 0. \tag{39}$$

The first term dominates as $\sigma \to 0$, so that $(\lambda^{[q]})^{\mathrm{T}}\tilde{\mathbf{u}} \geq (\lambda^{[q]})^{\mathrm{T}}\tilde{\mathbf{u}}_*^{[q]} = J_*^{[q]} > (1-\gamma)J_*^{[q]}$.

Since J is convex, it is supported by its linearization about $\tilde{\mathbf{u}}_\gamma^{[q]} \equiv (1-\gamma)\tilde{\mathbf{u}}_*^{[q]}$, and hence

$$J(\tilde{\mathbf{u}}_\gamma^{[q]}) + 2(1-\gamma)(\lambda^{[q]})^{\mathrm{T}}(\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_\gamma^{[q]}) \leq J(\tilde{\mathbf{u}}), \tag{40}$$

for all $\tilde{\mathbf{u}}$. If $\tilde{\mathbf{u}}$ violates (37), then the second term is positive, so that $J(\tilde{\mathbf{u}}) \geq J(\tilde{\mathbf{u}}_\gamma^{[q]}) = (1-\gamma)^2 J_*^{[q]}$.

Beginning with $q = 1$, the constraint $\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}} \notin \tilde{Z}^{[q]}$ in (19) is replaced with (37) if $J_*^{[q]} > 0$ (i.e., $\mathbf{0} \notin Z^{[q]}$). If (37) is applied for q, it is added to the definition of $\tilde{U}$ before (36) is solved for $q+1$, and this process is repeated until $q = Q$. By the first conclusion of Lemma 6, the end result is a restriction of (19). However, if the resulting program has an optimal objective value less than $(1-\gamma)^2 J_*^{[q]}$ for all q, then the second conclusion of Lemma 6 guarantees that the reformulated program is equivalent to the original. In general, this procedure is very useful when many scenarios are considered, since it is likely there are few, key pairs that are difficult to distinguish (i.e., those with $\mathbf{0} \in Z^{[q]}$), with the rest being relatively easy (i.e., those with $\mathbf{0} \notin Z^{[q]}$). This procedure is applied in Section 8 with $\gamma = 10^{-6}$.

6.2. Zonotope order reduction

Each constraint $\tilde{\mathbf{N}}^{[q]}\tilde{\mathbf{u}} \notin \tilde{Z}^{[q]}$ that is not eliminated through the methods of the previous section will contribute $(2(N+1)n_y\ell)$ binary variables to (33), where ℓ is the order of the zonotope $\tilde{Z}^{[q]}$. As discussed in Section 3.1, a given zonotope can be overapproximated by a zonotope of lower order (i.e., fewer generators) using standard techniques (Althoff et al., 2010). Thus, ℓ can be reduced by replacing $\tilde{Z}^{[q]}$ by such an overapproximation. Clearly, this approximation introduces some conservatism in (19). In particular, the optimal objective value may increase. Nonetheless, the optimal solution is guaranteed to be a separating input by Theorem 3. Moreover, reducing the order by one reduces the number of binary variables by a factor of $Q(2(N+1)n_y)$. In Section 8, it is shown that this approach can greatly simplify (33) with only a small impact on the optimal solution due to the added conservatism.

7. An observer-based diagnosis method

The previous section considered reductions in the number of binary variables $B = Q(2(N+1)n_y\ell)$ in (33) by reducing Q and ℓ. However, B still has an undesirable dependence on N, which may be large for some problems. This section presents a more conservative definition of a separating input in which the intersection condition (16) is replaced by the condition that the outputs of set-based observers (Combastel, 2003; Shamma, 1999) are disjoint at time N. Since this condition involves sets of dimension n_y rather than $(N+1)n_y$, such inputs can be computed through an analog of (33) with only $B = Q(2n_y\ell)$ binary variables. Furthermore, all of the data required to formulate this optimization are computed recursively (i.e., the high-dimensional sets in (16) need not be constructed), making this method tractable for large systems or large N.

For each $i \in \mathbb{I}$, choose $\mathbf{L}(i) \in \mathbb{R}^{n_x \times n_y}$ and define $\mathbf{A}_L(i) \equiv \mathbf{A}(i) - \mathbf{L}(i)\mathbf{C}(i)$. For each scenario $\tilde{i} \in \tilde{\mathbb{I}}$, we employ an observer of the form

$$\hat{\mathbf{x}}_{k+1} = \mathbf{A}(i_k)\hat{\mathbf{x}}_k + \mathbf{B}(i_k)\mathbf{u}_k + \mathbf{r}(i_k) + \mathbf{L}(i_k)(\mathbf{y}_k - \hat{\mathbf{y}}_k), \tag{41}$$

$$\hat{\mathbf{y}}_k = \mathbf{C}(i_k)\hat{\mathbf{x}}_k + \mathbf{s}(i_k), \tag{42}$$

with $\hat{\mathbf{x}}_0 = \mathbf{c}_0$ (recall the notation $X_0 = \{\mathbf{G}_0, \mathbf{c}_0\}$). Denote the state and output at k by $\hat{\phi}_k(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}})$ and $\hat{\psi}_k(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}})$, respectively. Regardless of $\tilde{\mathbf{u}}$, the errors $\hat{\mathbf{e}}_k \equiv \mathbf{x}_k - \hat{\mathbf{x}}_k$ and $\hat{\mathbf{f}}_k \equiv \mathbf{y}_k - \hat{\mathbf{y}}_k$ satisfy

$$\hat{\mathbf{e}}_{k+1} = \mathbf{A}_L(i_k)\hat{\mathbf{e}}_k + \mathbf{B}_w(i_k)\mathbf{w}_k - \mathbf{L}(i_k)\mathbf{D}_v(i_k)\mathbf{v}_k, \tag{43}$$

$$\hat{\mathbf{f}}_k = \mathbf{C}(i_k)\hat{\mathbf{e}}_k + \mathbf{D}_v(i_k)\mathbf{v}_k. \tag{44}$$

Define $E_0 \equiv X_0 - \mathbf{c}_0 = \{\mathbf{G}_0, \mathbf{0}\}$. Since $\hat{\mathbf{e}}_0 \in \hat{E}_0$, bounds on the errors $(\hat{\mathbf{e}}_k, \hat{\mathbf{f}}_k)$ can be computed recursively by

$$\hat{E}_{k+1} = \mathbf{A}_L(i_k)\hat{E}_k \oplus \mathbf{B}_w(i_k)W \oplus (-\mathbf{L}(i_k)\mathbf{D}_v(i_k))\,V, \tag{45}$$

$$\hat{F}_k = \mathbf{C}(i_k)\hat{E}_k \oplus \mathbf{D}_v(i_k)V. \tag{46}$$

For any $\tilde{\mathbf{u}} \in \mathbb{R}^{Nn_u}$, $\tilde{i} \in \tilde{\mathbb{I}}$, and $\tilde{\mathbf{y}}_{0:N-1} \in \mathbb{R}^{Nn_y}$, (41)–(42) and (45)–(46) together define a set-valued observer with state $\hat{\Phi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}) \equiv \hat{\phi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}) \oplus \hat{E}_N(\tilde{i})$ and output $\hat{\Psi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}) \equiv \hat{\psi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}) \oplus \hat{F}_N(\tilde{i})$. These sets can be computed recursively online. From the error bounds computed above, it is clear that the observer satisfies

$$\tilde{\mathbf{y}} \in \tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i}) \Longrightarrow \mathbf{y}_N \in \hat{\Psi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}). \tag{47}$$

Note, however, that the observer is not exact in the sense that the converse of (47) does not hold.

Now, consider a fault diagnosis scheme based on checking

$$\mathbf{y}_N \in \hat{\Psi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}) \tag{48}$$

online, for each $\tilde{i} \in \tilde{\mathbb{I}}$. If (48) fails for $\tilde{i} \in \tilde{\mathbb{I}}$, then the contrapositive of (47) implies that $\tilde{i}$ did not occur. This observation suggests the design of $\tilde{\mathbf{u}}$ such that (48) holds for exactly one $\tilde{i} \in \tilde{\mathbb{I}}$. However, it is only necessary to enforce this condition for certain $\tilde{\mathbf{y}} \in \mathbb{R}^{(N+1)n_y}$, as described in the following definition and subsequently discussed. The qualifier $\mathscr{L}$ below distinguishes this condition from (16) and reflects the dependence on $\mathbf{L}(i)$, $i \in \mathbb{I}$.

Definition 2. An input $\tilde{\mathbf{u}} \in \mathbb{R}^{Nn_u}$ is said to *$\mathscr{L}$-separate* $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ *at* N *given* $\tilde{i}$, if

$$\hat{\Psi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}) \cap \hat{\Psi}_N(\tilde{\mathbf{u}}, \tilde{j}, \tilde{\mathbf{y}}_{0:N-1}) = \emptyset, \tag{49}$$

$$\forall \tilde{\mathbf{y}}_{0:N-1} \in \tilde{\Psi}_{0:N-1}(\tilde{\mathbf{u}}_{0:N-2}, \tilde{i}_{0:N-1}). \tag{50}$$

Similarly, $\tilde{\mathbf{u}} \in \mathbb{R}^{Nn_u}$ is said to *$\mathscr{L}$-separate* $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ *at* N *given* $\tilde{j}$ if (49) holds $\forall \tilde{\mathbf{y}}_{0:N-1} \in \tilde{\Psi}_{0:N-1}(\tilde{\mathbf{u}}_{0:N-2}, \tilde{j}_{0:N-1})$. If $\tilde{\mathbf{u}}$ both $\mathscr{L}$-separates $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ at N given $\tilde{i}$ and given $\tilde{j}$, then it is simply said to *$\mathscr{L}$-separate* $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ *at* N. Finally, $\tilde{\mathbf{u}}$ *$\mathscr{L}$-separates* $\tilde{\mathbb{I}}$ *at* N, or is a *$\mathscr{L}$-separating input*, if it $\mathscr{L}$-separates every $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ with $\tilde{i} \neq \tilde{j}$.

To see that every $\mathscr{L}$-separating input guarantees diagnosis via the tests (48), suppose that $\tilde{\mathbf{u}}$ is injected, scenario $\tilde{i} \in \tilde{\mathbb{I}}$ occurs, and the output $\tilde{\mathbf{y}} \in \mathbb{R}^{(N+1)n_y}$ is measured. By (47), (48) holds for $\tilde{i}$. Moreover, $\tilde{\mathbf{y}}_{0:N-1}$ must satisfy (50). Thus, if $\tilde{\mathbf{u}}$ is a $\mathscr{L}$-separating input, then (49) and (50) imply that (48) fails for all other $\tilde{j} \in \tilde{\mathbb{I}}$. Therefore, checking (48) for all $\tilde{i} \in \tilde{\mathbb{I}}$ guarantees diagnosis at N. Note, however, that it is not required that all of the sets $\hat{\Psi}_N(\tilde{\mathbf{u}}, \tilde{j}, \tilde{\mathbf{y}}_{0:N-1})$ are mutually disjoint, but only that the observer set for the correct scenario is disjoint from all of the others, regardless of what the correct scenario turns out to be. Finally, when N is large, the inclusions (48) are much easier to check online than the test $\tilde{\mathbf{y}} \in \tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i})$ required to make use of separating inputs, because (48) involves an n_y-dimensional set that can be recursively computed, whereas $\tilde{\Psi}_{0:N}(\tilde{\mathbf{u}}, \tilde{i})$ is an $(N+1)n_y$-dimensional set and cannot be recursively computed (see Section 3.2).

Although $\mathscr{L}$-separating inputs have an intuitive interpretation, a weaker condition suffices to ensure diagnosis via (48), and this condition enhances the computational advantages of the observer-based method, as discussed in the next section.

Definition 3. An input $\tilde{\mathbf{u}} \in \mathbb{R}^{Nn_u}$ is said to *$\mathscr{L}^*$-separate* $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ *at* N if it either $\mathscr{L}$-separates $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ at N given $\tilde{i}$ or given $\tilde{j}$, but not necessarily both. Similarly, $\tilde{\mathbf{u}}$ *$\mathscr{L}^*$-separates* $\tilde{\mathbb{I}}$ *at* N, or is a *$\mathscr{L}^*$-separating input*, if it $\mathscr{L}^*$-separates every $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$ with $\tilde{i} \neq \tilde{j}$.

If $\tilde{\mathbf{u}}$ is a $\mathscr{L}^*$-separating input, then it can happen that $\mathbf{y}_N \in \hat{\Psi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}) \cap \hat{\Psi}_N(\tilde{\mathbf{u}}, \tilde{j}, \tilde{\mathbf{y}}_{0:N-1})$ for some $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$. Nonetheless, the tests (48) still suffice to determine the correct scenario. For example, if $\tilde{\mathbf{u}}$ was designed to separate $\tilde{i}, \tilde{j}$ at N given $\tilde{i}$ (as opposed to given $\tilde{j}$), then the failure of (49) immediately implies that $\tilde{\mathbf{y}}_{0:N-1} \notin \tilde{\Psi}_{0:N-1}(\tilde{\mathbf{u}}_{0:N-2}, \tilde{i}_{0:N-1})$, and hence $\tilde{i}$ did not occur. This reasoning can be used to eliminate one scenario from every pair with $\mathbf{y}_N$ in the intersection of their observer output sets, leaving only the correct scenario. Despite the need for this additional reasoning, optimal $\mathscr{L}^*$-separating inputs are easier to compute than optimal $\mathscr{L}$-separating inputs (see Section 7.1), and can clearly achieve lower objective values.

Any $\mathscr{L}^*$-separating input (and hence any $\mathscr{L}$-separating input) is also a separating input in the sense of Definition 1, but the converse is not true because the observer sets $\hat{\Psi}_N$ are approximate (i.e., the converse of (47) is false). The proofs of these assertions

are straightforward and are omitted for brevity. The practical consequence is that optimal separating inputs are guaranteed to be at least as good as optimal $\mathscr{L}^*$-separating inputs. On the other hand, $\mathscr{L}^*$-separating inputs are much easier to compute when N is large.

7.1. Optimal $\mathscr{L}$- and $\mathscr{L}^*$-separating inputs

This section shows that the set of $\mathscr{L}^*$-separating inputs can be described in the form

$$\{\tilde{\mathbf{u}} \in \mathbb{R}^{Nn_u} : \mathbf{N}^{[q]}\tilde{\mathbf{u}} \notin Z^{[q]},\ q = 1, \dots, Q\}, \tag{51}$$

for matrices $\mathbf{N}^{[q]} \in \mathbb{R}^{n_y \times Nn_u}$ and zonotopes $Z^{[q]} \subset \mathbb{R}^{n_y}$ to be derived. Of course, (51) is exactly the form used for the set of separating inputs in Section 5, so that the optimization procedure described there can be applied directly to compute optimal $\mathscr{L}^*$-separating inputs.

Let $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}, \tilde{i} \neq \tilde{j}$. It suffices to derive a constraint $\mathbf{N}^{[q]}\tilde{\mathbf{u}} \notin Z^{[q]}$ equivalent to the condition that $\tilde{\mathbf{u}}$ $\mathscr{L}$-separates $\tilde{i}, \tilde{j}$ at N given $\tilde{i}$. Adding one such constraint for each distinct pair of scenarios specifies the set of $\mathscr{L}^*$-separating inputs, and the $\mathscr{L}$-separating inputs can be specified by adding a second constraint for each pair that is derived analogously.

To begin, (49) is reformulated using Lemma 2 to obtain

$$\hat{\psi}_N(\tilde{\mathbf{u}}, \tilde{j}, \tilde{\mathbf{y}}_{0:N-1}) - \hat{\psi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1}) \notin \hat{F}_N(\tilde{i}) \oplus \hat{F}_N(\tilde{j}). \tag{52}$$

The set of differences $\hat{\psi}_N(\tilde{\mathbf{u}}, \tilde{j}, \tilde{\mathbf{y}}_{0:N-1}) - \hat{\psi}_N(\tilde{\mathbf{u}}, \tilde{i}, \tilde{\mathbf{y}}_{0:N-1})$ achievable with $\tilde{\mathbf{y}}_{0:N-1}$ generated by scenario $\tilde{i}$ can be computed recursively as follows. Assuming that $\tilde{i}$ occurs, consider the evolution of $\mathbf{q}_k = (\hat{\mathbf{x}}_k(\tilde{i}), \hat{\mathbf{x}}_k(\tilde{j}), \mathbf{x}_k)$, where $\mathbf{x}_k$ is the system state, and $\hat{\mathbf{x}}_k(\tilde{i})$ and $\hat{\mathbf{x}}_k(\tilde{j})$ are the states of the observers for scenarios $\tilde{i}$ and $\tilde{j}$, respectively. Straightforward algebra shows that $\mathbf{q}_k$ and the corresponding output $\mathbf{p}_k = (\hat{\mathbf{y}}_k(\tilde{i}), \hat{\mathbf{y}}_k(\tilde{j}), \mathbf{y}_k)$ obey

$$\mathbf{q}_{k+1} = \mathscr{A}_k\mathbf{q}_k + \mathscr{B}_k\mathbf{u}_k + \mathscr{B}_{w,k}\mathbf{w}_k + \mathscr{B}_{v,k}\mathbf{v}_k + \mathscr{R}_k, \tag{53}$$

$$\mathbf{p}_k = \mathscr{C}_k\mathbf{q}_k + \mathscr{D}_{v,k}\mathbf{v}_k + \mathscr{S}_k, \tag{54}$$

where $\mathscr{B}_k \equiv \left[\mathbf{B}^{\mathrm{T}}(i_k)\ \mathbf{B}^{\mathrm{T}}(j_k)\ \mathbf{B}^{\mathrm{T}}(i_k)\right]^{\mathrm{T}}$, $\mathscr{B}_{w,k} \equiv \left[\mathbf{0}\ \mathbf{0}\ \mathbf{B}_w^{\mathrm{T}}(i_k)\right]^{\mathrm{T}}$, $\mathscr{D}_{v,k} \equiv \left[\mathbf{0}\ \mathbf{0}\ \mathbf{D}_v^{\mathrm{T}}(i_k)\right]^{\mathrm{T}}$, $\mathscr{B}_{v,k} \equiv \left[(\mathbf{L}(i_k)\mathbf{D}_v(i_k))^{\mathrm{T}}\ (\mathbf{L}(j_k)\mathbf{D}_v(i_k))^{\mathrm{T}}\ \mathbf{0}\right]^{\mathrm{T}}$, $\mathscr{S}_k \equiv \big(\mathbf{s}(i_k), \mathbf{s}(j_k), \mathbf{s}(i_k)\big)$, $\mathscr{R}_k \equiv \big(\mathbf{r}(i_k), \mathbf{r}(j_k) + \mathbf{L}(j_k)(\mathbf{s}(i_k) - \mathbf{s}(j_k)), \mathbf{r}(i_k)\big)$,

$$\mathscr{A}_k \equiv \begin{bmatrix} \mathbf{A}_L(i_k) & \mathbf{0} & \mathbf{L}(i_k)\mathbf{C}(i_k) \\ \mathbf{0} & \mathbf{A}_L(j_k) & \mathbf{L}(j_k)\mathbf{C}(j_k) \\ \mathbf{0} & \mathbf{0} & \mathbf{A}(i_k) \end{bmatrix},$$

$$\mathscr{C}_k \equiv \begin{bmatrix} \mathbf{C}(i_k) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}(j_k) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}(i_k) \end{bmatrix}.$$

Note that (53)–(54) is a linear system with zonotopic errors and zonotopic initial condition set $\{\mathbf{c}_0\} \times \{\mathbf{c}_0\} \times X_0$. From the derivations in Section 3.2, it follows that the set of all possible $\mathbf{p}_N$ is a zonotope of the form $\Gamma_N(\tilde{\mathbf{u}}) = \{\mathbf{G}_N^{\Gamma}, \overline{\mathscr{B}}_N^{\Gamma}\tilde{\mathbf{u}} + \overline{\gamma}_N\}$, where $\overline{\gamma}_N$ is the output of (53)–(54) at N when $\mathbf{x}_0 = \mathbf{c}_0$, $(\mathbf{w}_k, \mathbf{v}_k) = (\mathbf{c}_W, \mathbf{c}_V)$, and $\mathbf{u}_k = \mathbf{0}$, $\forall k$. Moreover, $\mathbf{G}_N^{\Gamma}$, $\overline{\mathscr{B}}_N^{\Gamma}$, and $\overline{\gamma}_N$ can all be computed recursively as described in Section 3.2.

From the definition of $\mathbf{p}_N$, the set of differences on the left-hand side of (52) is given by $[-\mathbf{I}\ \mathbf{I}\ \mathbf{0}]\Gamma_N(\tilde{\mathbf{u}})$. Thus, (52) becomes $[-\mathbf{I}\ \mathbf{I}\ \mathbf{0}]\Gamma_N(\tilde{\mathbf{u}}) \cap (\hat{F}_N(\tilde{i}) \oplus \hat{F}_N(\tilde{j})) = \emptyset$. By a final application of Lemma 2, this is equivalent to

$$[\mathbf{I}\ -\mathbf{I}\ \mathbf{0}]\overline{\mathscr{B}}_N^{\Gamma}\tilde{\mathbf{u}} \notin \hat{F}_N(\tilde{i}) \oplus \hat{F}_N(\tilde{j}) \oplus [\mathbf{I}\ -\mathbf{I}\ \mathbf{0}]\Gamma_N(\mathbf{0}), \tag{55}$$

using the fact that $\mathbf{M}\{\mathbf{G}, -\mathbf{c}\} = \mathbf{M}\{-\mathbf{G}, -\mathbf{c}\} = -\mathbf{M}\{\mathbf{G}, \mathbf{c}\}$ for any matrix $\mathbf{M}$ and zonotope $\{\mathbf{G}, \mathbf{c}\}$.

Eq. (55) is now in the desired form, so that an optimal $\mathscr{L}$- or $\mathscr{L}^*$-separating input can be computed exactly as in Section 5. However, the zonotope on the right-hand side of (55) is n_y-dimensional as opposed to $(N+1)n_y$-dimensional. With fixed zonotope order ℓ, such a set is described by $n_y\ell$ generators, leading to $Q(2n_y\ell)$ binary variables in the computation of an optimal $\mathscr{L}^*$-separating input (this increases to $(2Q)(2n_y\ell)$ for $\mathscr{L}$-separating inputs because two constraints are required for each pair $\tilde{i}, \tilde{j} \in \tilde{\mathbb{I}}$). Clearly, the resulting reduction in computational time can be very substantial when N is large. On the other hand, $\mathscr{L}^*$-separating inputs are more conservative than separating inputs due to the conservatism of the set-based observers, so separating inputs are preferred when the computational cost is manageable.

7.2. Choosing the observer gains

The choice of the matrices $\mathbf{L}(i)$, $i \in \mathbb{I}$, has a large impact on the performance of the observer-based method. As a limiting case, it is simple to show that choosing $\mathbf{L}(i) = \mathbf{0}$, $\forall i \in \mathbb{I}$, gives the condition for separation

$$\Psi_N(\tilde{\mathbf{u}}, \tilde{i}) \cap \Psi_N(\tilde{\mathbf{u}}, \tilde{j}) = \emptyset. \tag{56}$$

In words, this choice requires the reachable sets at the terminal time N to be disjoint, so that diagnosis can be achieved solely on the basis of the terminal measurement $\mathbf{y}_N$. This choice performs reasonably well in practice, especially when $\mathrm{rank}(\mathbf{C}(i)) = n_x$ for all models. In contrast to non-zero choices of $\mathbf{L}(i)$, note that the sets in (56) do not depend on the output measurements, and that (56) is symmetric with respect to interchange of $\tilde{i}$ and $\tilde{j}$, so that $\mathscr{L}$- and $\mathscr{L}^*$-separating inputs coincide.

Assuming that each model is observable, nontrivial observer matrices can be designed so that each $\mathbf{A}_L(i)$ is stable using standard algorithms, such as Kalman filtering or pole placement. The latter choice ensures that the error sets $\hat{E}_k$ are described by stable dynamics and works well for most cases attempted. However, the results are highly dependent on the user-specified pole locations. At present, it is not clear how to choose optimal observer matrices in the context of the proposed diagnosis scheme. We leave this topic for future research.

8. Numerical examples

Consider the second-order linear nominal system

$$\mathbf{A}(1) = \begin{bmatrix} 0.6 & 0.2 \\ -0.4 & -0.2 \end{bmatrix}, \quad \mathbf{B}(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\mathbf{C}(1) = \begin{bmatrix} 1 & 0 \end{bmatrix},$$

$$\mathbf{B}_w(1) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{D}_v(1) = \begin{bmatrix} 1 \end{bmatrix}, \quad \mathbf{r}(1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\mathbf{s}(1) = \begin{bmatrix} 0 \end{bmatrix},$$

with four fault models, $i = 2, \dots, 5$, defined by the modifications to the nominal model:

$$\mathbf{B}(2) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B}(3) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\mathbf{A}(4) = \begin{bmatrix} 1.2 & 0.2 \\ -0.4 & -0.2 \end{bmatrix}, \quad \mathbf{A}(5) = \begin{bmatrix} 2.0 & 0.2 \\ -0.4 & -0.7 \end{bmatrix}.$$

Models 2 and 3 have faulty actuators, and models 4 and 5 represent system faults. In generator notation, define $X_0 = \{0.2\mathbf{I}, (-3, -3)\}$, $W = \{0.5\mathbf{I}, \mathbf{0}\}$ and $V = \{0.2, \mathbf{0}\}$.

First, the method of Section 5 was applied to synthesize an input that distinguishes between models $\{1, \dots, 5\}$ in a minimum number of steps N. For convenience, the method of Section 5 will be referred to as the full-measurement method below. It is assumed that one model is active on all of $[0, N]$; i.e., $\tilde{\mathbb{I}}$ contains the five fault

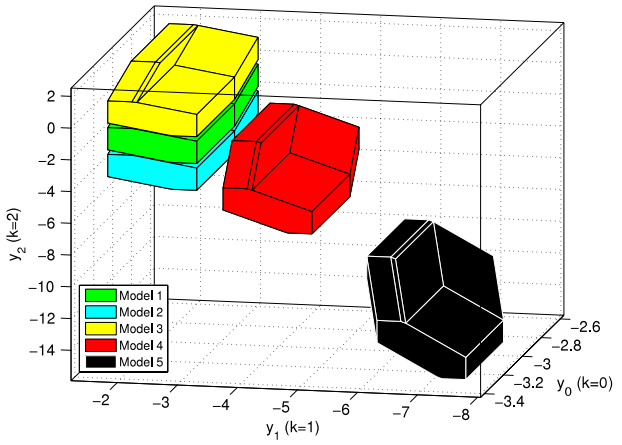

Fig. 1. *Full-measurement method*: Reachable output sets on [0, 2] given the optimal separating input with $N = 2$ and $\ell = 2$.

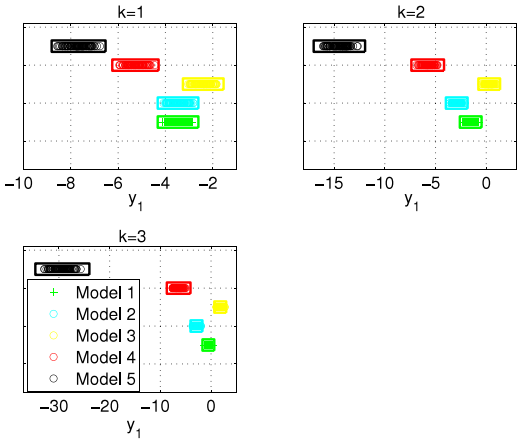

Fig. 2. *Observer-based method*, $\mathbf{L}(i) = \mathbf{0}$: One-dimensional observer output sets (boxes) at $k = 1, 2$, and 3 given the optimal $\mathscr{L}^*$-separating input with $N = 3$ and $\ell = 1$. Sets are plotted at different vertical positions for clarity. Circles represent 1000 output samples from the nominal and faulty systems.

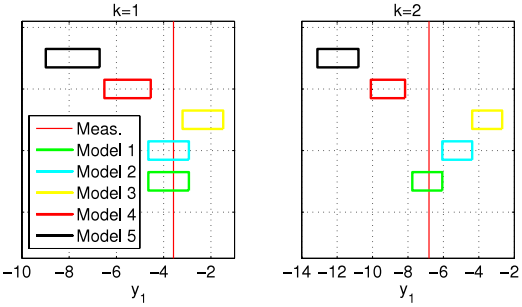

Fig. 3. *Observer-based method, Kalman observer gain*: One-dimensional observer output sets (boxes) at $k = 1, 2$, and 3, given the optimal $\mathscr{L}$-separating input with $N = 2$ and $\ell = 1$, for one simulation of the system with $i = 1$. Sets are plotted at different vertical positions for clarity. The red line represents the measurement.

scenarios $(1, \ldots, 1), \ldots, (5, \ldots, 5)$. For optimization, $\mathbf{R} = \mathbf{I}$ and $U = \{\mathbf{u} : \|\mathbf{u}\|_\infty \leq 9\}$. To limit the number of binary variables in (33), the zonotope order is limited to $\ell = 1$ or $\ell = 2$ below, resulting in $2Q(N+1)n_y\ell = 10(N+1)$ or $20(N+1)$ binary variables. The method was feasible for $N \geq 2$. With $\ell = 1$ and $N = 2$, the minimum two-norm separating input was found in 0.02 s,[2] with $\|\tilde{\mathbf{u}}\|_2 = 8.6143$. Increasing ℓ to 2, the method required 3.58 s and found a minimum norm of 8.6081, which is a slight reduction at the cost of a significant increase in the computational time. These results did not make use of the pair elimination method of Section 6.1. Including this procedure, 3 pairs are eliminated based on infeasibility of (36) (i.e., $\nexists \tilde{\mathbf{u}} \in \tilde{U}$ such that $\tilde{\mathbf{N}}\tilde{\mathbf{u}} \in \tilde{Z}^{[q]}$). With $\ell = 1$, an additional pair is eliminated through the inclusion of a linear constraint leaving $Q' = 6$ pairs, and with $\ell = 2$, four additional pairs are eliminated in this way, leaving $Q' = 3$. The drastic reduction when $\ell = 2$ results in a speed-up of one order of magnitude (0.36 s) while providing the same optimal objective value as the original problem ($\|\tilde{\mathbf{u}}\|_2 = 8.6081$). Fig. 1 shows the reachable output sets on [0, 2] for all models given the optimal separating input. Clearly, these sets are disjoint, so that output measurements on [0, 2] will be consistent with at most one fault scenario, providing the desired fault diagnosis.

Next, the observer-based method of Section 7 is applied to the same problem and compared. Two choices of the observer gain for each $i \in \mathbb{I}$ are considered: (a) $\mathbf{L}(i) = \mathbf{0}$, and (b) $\mathbf{L}(i)$ is the Kalman observer gain computed with weighting matrices $\mathbf{Q}' = 1000\mathbf{I}$ and $\mathbf{R}' = \mathbf{I}$ (not to be confused with $\mathbf{R}$ used to define J in Section 5). Choice (a) will be referred to as the terminal-measurement method following the discussion in Section 7.2. The advantage of the full-measurement method over the terminal-measurement method can be seen in Fig. 1. Although the 3-dimensional sets in Fig. 1 are disjoint, their projections onto the y_2-axis are not. Therefore, the optimal separating input is not $\mathscr{L}^*$-separating for choice (a). The terminal method is found to be feasible for $N \geq 3$. With $N = 3$ and $\ell = 1$, the minimum two-norm $\mathscr{L}^*$-separating input was found in 0.01 s, with $\|\tilde{\mathbf{u}}\|_2 = 11.5203$ (see Fig. 2). Using choice (b), $\mathscr{L}^*$-separating inputs exist for $N \geq 2$. With $N = 2$ and $\ell = 1$, the minimum two-norm $\mathscr{L}^*$-separating input was found in 0.01 s, with $\|\tilde{\mathbf{u}}\|_2 = 9.2144$. This example shows that non-zero observer gains can provide a reduction in input length and norm. Fig. 3 shows the results for choice (b). Note that, when $\mathbf{L}(i) \neq \mathbf{0}$, the position of the output reachable sets depends on the measurements. Fig. 3 shows a single simulation assuming model $i = 1$ is correct. According to the definition of an $\mathscr{L}^*$-separating input, it is guaranteed for such a simulation that the correct model can be determined via the tests (48), as described in Section 7. Under the same input, an analogous guarantee holds regardless of the correct model. However, the observer sets need not be mutually disjoint as they are in Fig. 3. For this example, increasing ℓ did not provide a benefit for either observer-based approach.

To illustrate the computational aspects of the proposed approaches, they were applied to randomly generated models with $n_x = 50$ and $n_y = n_u = n_w = n_v = 10$. The MATLAB™ function *drss* was used to generate systems with random stable poles (with the possible exception of poles at $z = 1$). First-order zonotopes X_0, W, and V were also randomly generated. The input constraints $U = \{\mathbf{u} : \|\mathbf{u}\|_\infty \leq 20\}$ were fixed for all tests. The full-measurement method and observer-based method (with $\mathbf{L}(i) = \mathbf{0}$) were first evaluated for the separation of two randomly generated systems with $N = 10$ and $\ell = 1$. Over 100 tests, the average computation time of the full method was 1.13 s, compared to 0.04 s for the observer-based method. In terms of norm, the full method performed much better (19.01 vs. 41.70). Increasing the number of models to 3, the average time for the full method increased to 50.30 s, compared to 1.04 s for the observer-based method, which clearly shows the increase in computational complexity for additional models. Again, the minimum norm obtained with the full method was considerably smaller (247.10 vs. 573.99). When considering longer horizons, the full method became too costly due to the dependence of the number of binary variables on N. On the other hand, the observer-based method was able to solve problem instances with $N = 100$ in an average time of 20.42 s. The number of binary variables required for this method is independent of N. Compared to the case where $N = 10$, the mild increase

[2] Laptop (Intel i7, 2.67 GHz, 4 GB RAM) running Windows 7 and using a single core; optimization using CPLEX 12.2 (IBM, 2012).

Table 1
Description of faults.

Model	Fault type
1	Fault-free
2	Increase of armature resistance (+0.5 Ω)
3	Increase of armature resistance (+1.14 Ω)
4	Wearing of brush, insufficient brush pressure
5	Short circuit of two commutator bars
6	Disconnection of coil from commutator bar

in time is due to the increased number of continuous decision variables $\tilde{\mathbf{u}}$. Considering two models with increased zonotope order $\ell = 2$, average times for $N = 10$ and $N = 100$ were found to be 6.0358 s and 18.4609 s, respectively.

8.1. Permanent-magnet DC motor

A low-frequency linear model for a permanent-magnet DC motor can be developed from basic physical laws as:

$$\begin{bmatrix} \frac{di(t)}{dt} \\ \frac{dn(t)}{dt} \end{bmatrix} = \begin{bmatrix} -R_a/L & -K_e/L \\ K_t/J_1 & -f_r/J_1 \end{bmatrix} \begin{bmatrix} i(t) \\ n(t) \end{bmatrix} + \begin{bmatrix} 1/L \\ 0 \end{bmatrix} u(t)$$

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix},$$

where u, i, R_a, L, K_e, K_t, J_1, and f_r denote the armature voltage, current, resistance, inductance, torque constant, back EMF constant, motor inertia, and friction coefficient, respectively. Six fault models are described in Table 1, corresponding to the model parameters in Table 2 (Liu, Zhang, Liu, & Yang, 2000). For each fault, the parameters in Table 2 were identified from experimental data.

The torque constant K_t (N m/amp) is related to K_e by $K_t = 1.0005K_e$. To keep the nominal speed of the motor at around 70.3 rad/s, the nominal input was set to $u_c = 6$ V, which was modeled by adding $\mathbf{r}(i) = \mathbf{B}(i)u_c$ to the state equations. The active input u_s was added to u_c, subject to the constraint $|u_s| \leq 6$ V. The models were discretized by forward Euler differencing with a sampling interval of 5 ms. The initial condition and measurement errors lie in $X_0 = \left\{ \begin{bmatrix} 0.06 & 0 \\ 0 & 0.6 \end{bmatrix}, \begin{bmatrix} 0.6 \\ 70 \end{bmatrix} \right\}$ and $V = \left\{ \begin{bmatrix} 0.06 & 0 \\ 0 & 0.6 \end{bmatrix}, \mathbf{0} \right\}$, respectively. In order to account for parameter uncertainties, the discretized state dynamics were augmented with the term $\mathbf{B}_w(i)\mathbf{w}_k$, where

$$\mathbf{B}_w(1) = \begin{bmatrix} -0.0254 & -0.0778 \\ -0.3996 & 0.3026 \end{bmatrix},$$

$$\mathbf{B}_w(2) = \begin{bmatrix} -0.0231 & -0.0471 \\ -0.3470 & 0.2798 \end{bmatrix},$$

$$\mathbf{B}_w(3) = \begin{bmatrix} -0.0227 & -0.0346 \\ -0.3113 & 0.2230 \end{bmatrix},$$

$$\mathbf{B}_w(4) = \begin{bmatrix} -0.0242 & -0.0537 \\ -0.3516 & 0.2797 \end{bmatrix},$$

$$\mathbf{B}_w(5) = \begin{bmatrix} -0.0241 & -0.0661 \\ -0.3672 & 0.3154 \end{bmatrix},$$

$$\mathbf{B}_w(6) = \begin{bmatrix} -0.0282 & -0.0589 \\ -0.3926 & 0.1684 \end{bmatrix},$$

and $\mathbf{w}_k$ lies in $W = \{\mathbf{I}, \mathbf{0}\}$. These matrices were obtained by assuming 5% uncertainty in R_a, K_e, J_1, and f_r, and computing the worst-case additive error with the current and motor speed bounded in $[-2, 2]$ A and $[-150, 150]$ rad/s, respectively.

The problem of separating the six models in Table 1 was considered under the assumption that one model is active on the whole interval of interest. That is, $\mathbb{I}$ consists of the six scenarios $(1, \ldots, 1), \ldots, (6, \ldots, 6)$. The observer-based method with $\mathbf{L}(i) = \mathbf{0}$ was applied with zonotope order reduction to $\ell = 3$. The minimum horizon N for which a feasible $\mathscr{L}^*$-separating input exists was found to be 9. The minimum two-norm separating input was found in 58.47 s, with a norm of 15.92 (see Fig. 4). In this case, applying the pair elimination approach of Section 6.1 reduces the number of pairs from 15 to 10, but does not preserve the optimal solution of the original problem; the minimum horizon increases to 10 with optimal input norm $\|\tilde{\mathbf{u}}\| = 16.0095$. On the other hand, the computational time is reduced dramatically, to only 0.065 s.

Table 2
Fault model parameters.

Model	R_a (Ω)	L (10^{-3} H)	K_e ($10^{-2}\ \frac{\text{V rad}}{\text{s}}$)	J_1 ($10^{-4}\ \frac{\text{J s}^2}{\text{rad}}$)	f_r ($10^{-4}\ \frac{\text{J s}}{\text{rad}}$)
1	1.2030	5.5840	8.1876	1.3528	2.3396
2	1.7725	5.5837	8.0203	1.3320	2.3769
3	2.2837	6.4942	8.1094	1.4503	1.9291
4	1.7690	6.0798	8.7987	1.4964	2.3570
5	1.1743	4.4053	7.0094	1.1664	3.8335
6	1.4365	8.7548	7.7020	1.4185	4.1279

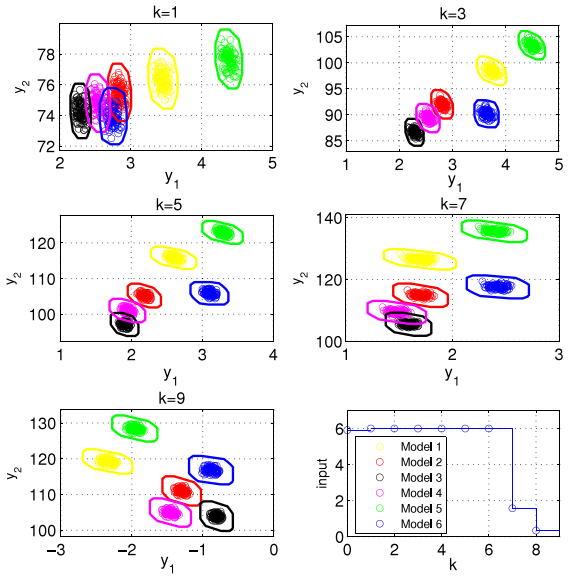

Fig. 4. Optimal $\mathscr{L}^*$-separating input with $\mathbf{L}(i) = \mathbf{0}$ (lower right) separating observer output sets for six DC motor fault models on the interval [0, 9] (upper left–lower left). Circles represent 1000 samples of the nominal and faulty outputs.

Considering the 3 scenarios $(2, \ldots, 2), (4, \ldots, 4), (6, \ldots, 6)$, the results of the observer-based method were compared with $\mathbf{L}(i) = \mathbf{0}$ and $\mathbf{L}(i)$ designed via pole placement (the Kalman observer design did not provide satisfactory results in this case). In the second case, $\mathbf{L}(i)$ was chosen to place the eigenvalues of $\mathbf{A}_L(i)$ at zero for all $i = 2, 4, 6$. Table 3 reports the results for different values of ℓ, which clearly show that an appropriate design of $\mathbf{L}(i)$ can drastically improve performance. The full-measurement method for this example was feasible for $N \geq 2$. With $N = 2$ and $\ell = 1$ the minimum two-norm separating input was found in 0.035 s, with a norm of 6.5762.

9. Conclusions

A deterministic method was proposed for computing the set of inputs that are guaranteed to lead to a fault diagnosis in a specified period of time, provided that such inputs exist. This set is shown to be characterized in terms of the complement of a finite number of zonotopes, and can be computed efficiently and reliably. Furthermore, a computationally practical optimization formulation has been derived for choosing an optimal separating

Table 3
Effect of gain on observer-based method.

ℓ	$\mathbf{L}(i) = \mathbf{0}$			Pole placement		
	N	$\|\tilde{\mathbf{u}}\|$	Time	N	$\|\tilde{\mathbf{u}}\|$	Time
1	7	13.571	0.02	2	7.182	0.01
2	7	13.571	0.02	2	7.182	0.01
3	7	13.571	0.07	2	7.182	0.03
4	7	13.559	0.28	2	6.846	0.71
5	3	8.433	4.41	2	6.846	4.33
6	3	8.230	36.71	2	6.846	22.35

input in a flexible way. Using this formulation, a separating input of minimum norm can be computed, or the set of separating inputs can be used within a more complex predictive control calculation, which will be the focus of subsequent investigations.

Acknowledgment

The authors would like to thank Roberto Marseglia from the Identification and Control of Dynamic Systems Laboratory, University of Pavia, for assistance with simulations.

References

Althoff, M., & Krogh, B. H. (2011). Zonotope bundles for the efficient computation of reachable sets. In *Proc. 50th IEEE conference on decision and control* (pp. 6814–6821).
Althoff, M., Stursberg, O., & Buss, M. (2010). Computing reachable sets of hybrid systems using a combination of zonotopes and polytopes. *Nonlinear Analysis—Hybrid Systems*, *4*(2), 233–249.
Andjelkovic, I., Sweetingham, K., & Campbell, S. L. (2008). Active fault detection in nonlinear systems using auxiliary signals. In *Proc. of the American control conference* (pp. 2142–2147).
Ashari, A. E., Nikoukhah, R., & Campbell, S. L. (2009). Asymptotic behavior and solution approximation of active robust fault detection for closed-loop systems. In *Proc. of the 48th IEEE conference on decision and control* (pp. 1026–1031).
Ashari, A., Nikoukhah, R., & Campbell, S. L. (2012). Effects of feedback on active fault detection. *Automatica*, *48*, 866–872.
Bertsekas, D. P. (1999). *Nonlinear programming* (2nd ed.). Belmont, MA: Athena Scientific.
Bertsimas, D., & Tsitsiklis, J. (1997). *Introduction to linear optimization*. Belmont, MA: Athena Scientific.
Blackmore, L., Rajamanoharan, S., & Williams, B. C. (2008). Active estimation for jump Markov linear systems. *IEEE Transactions on Automatic Control*, *53*(10), 2223–2236.
Bolker, E. D. (1969). A class of convex bodies. *Transactions of the American Mathematical Society*, *145*, 323–345.
Campbell, S. L., Drake, K., & Nikoukhah, R. (2002). Early decision making when using proper auxiliary signals. In *Proc. of the 41st IEEE conference on decision and control*. December (pp. 1832–1837).
Campbell, S. L., Horton, K. G., & Nikoukhah, R. (2002). Auxiliary signal design for rapid multi-model identification using optimization. *Automatica*, *38*(8), 1313–1325.
Campbell, S. L., & Nikoukhah, R. (2004). *Auxiliary signal design for failure detection*. Princeton University Press.
Chiang, L. H., Russell, E. L., & Braatz, R. D. (2001). *Fault detection and diagnosis in industrial systems*. London: Springer-Verlag.
Combastel, C. (2003). A state bounding observer based on zonotopes. In *Proc. of the European control conference*. Cambridge, UK.
Dobkin, D., Hershberger, J., Kirkpatrick, D., & Suri, S. (1993). Computing the intersection-depth of polyhedra. *Algorithmica*, *9*, 518–533.
Esna Ashari, A., Nikoukhah, R., & Campbell, S. L. (2012). Active robust fault detection in closed-loop systems: quadratic optimization approach. *IEEE Transactions on Automatic Control*, *57*(10), 2532–2544.
Fortuny-Amat, J., & McCarl, B. (1981). A representation and economic interpretation of a two-level programing problem. *Journal of the Operational Research Society*, *32*, 783–792.
Fukuda, K. (2004). From the zonotope construction to the Minkowski addition of convex polytopes. *Journal of Symbolic Computation*, *38*(4), 1261–1272.
IBM (2012). ILOG CPLEX V12.2 user's manual for CPLEX.
Kolmanovsky, I., & Gilbert, E. G. (1998). Theory and computation of disturbance invariant sets for discrete-time linear systems. *Mathematical Problems in Engineering*, *4*, 317–367.
Kuhn, W. (1998). Rigorously computed orbits of dynamical systems without the wrapping effect. *Computing*, *61*(1), 47–67.
Liu, X. Q., Zhang, H. Y., Liu, J., & Yang, J. (2000). Fault detection and diagnosis of permanent-magnet DC motor based on parameter estimation and neural network. *IEEE Transactions on Industrial Electronics*, *47*(5), 1021–1030.
Niemann, H. H. (2006). A setup for active fault diagnosis. *IEEE Transactions on Automatic Control*, *51*, 1572–1578.
Nikoukhah, R. (1998). Guaranteed active failure detection and isolation for linear dynamical systems. *Automatica*, *34*(11), 1345–1358.
Nikoukhah, R., & Campbell, S. L. (2006). Auxiliary signal design for active failure detection in uncertain linear systems with a priori information. *Automatica*, *42*(2), 219–228.
Nikoukhah, R., Campbell, S. L., & Delebecque, F. (2000). Detection signal design for failure detection: a robust approach. *International Journal of Adaptive Control and Signal Processing*, *14*(7), 701–724.
Scott, J. K., Findeisen, R., Braatz, R. D., & Raimondo, D. M. (2013). Design of active inputs for set-based fault diagnosis. In *Proc. of the American control conference* (pp. 3567–3572).
Shamma, J. S. (1999). Set-valued observers and optimal disturbance rejection. *IEEE Transactions on Automatic Control*, *44*, 253–264.
Simandl, M., & Puncochar, I. (2009). Active fault detection and control: unified formulation and optimal design. *Automatica*, *45*(9), 2052–2059.
Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. (2003). A review of process fault detection and diagnosis: part I: quantitative model-based methods. *Computers & Chemical Engineering*, *27*(3), 293–311.
Zolghadri, A. (2010). Advanced model-based FDIR techniques for aerospace systems: today challenges and opportunities. *Progress in Aerospace Sciences*, *53*(3), 18–29.

Joseph K. Scott is an assistant professor in the Department of Chemical and Biomolecular Engineering at Clemson University. He was born in 1984 in Royal Oak, MI, USA. He received his B.S. (2006) in Chemical Engineering from Wayne State University, and his M.S. (2008) and Ph.D. (2012) in Chemical Engineering from the Massachusetts Institute of Technology. His research interests include dynamical systems, differential-algebraic equations, dynamic optimization, global optimization, reachability analysis, and renewable energy systems.

Rolf Findeisen graduated with an M.S. degree from the University of Wisconsin, Madison and a Diploma and Doctorate from the University of Stuttgart, Germany. He was a research assistant in the Automatic Control Laboratory at ETH Zürich and a researcher at the Institute for Systems Theory and Automatic Control at the University of Stuttgart. He currently is a full Professor and head of the Systems Theory and Automatic Control Laboratory at the Otto-von-Guericke University, Magdeburg, Germany. His research focuses on the method and theory development for the analysis and control of complex nonlinear systems. The main interest is in optimization based approaches, predictive control, set-based estimation and validation methods, process automation and various fields of applications, spanning from mechatronics to systems biology.

Richard D. Braatz is the Edwin R. Gilliland Professor at the Massachusetts Institute of Technology (MIT) where he does research in applied mathematics and control theory and its application to manufacturing processes, biomedical systems, and nanotechnology. He received M.S. and Ph.D. degrees from the California Institute of Technology and was on the faculty at the University of Illinois at Urbana-Champaign and was a Visiting Scholar at Harvard University before moving to MIT. He has consulted or collaborated with more than 20 companies including United Technologies Corporation, IBM, BP, and Novartis. Honors include the AACC Donald P. Eckman Award, the Antonio Ruberti Young Researcher Prize, the IEEE Control Systems Society Transition to Practice Award, and best paper awards from IEEE- and IFAC-sponsored control journals. He is a Fellow of IEEE and IFAC.

Davide M. Raimondo was born in Pavia, Italy, in 1981. He received the B.Sc. and M.Sc. in Computer Engineering, and the Ph.D. in Electronic, Computer Science and Electric Engineering from the University of Pavia, Italy, in 2003, 2005, and 2009, respectively. As a Ph.D. student he held a visiting position at the Department of Automation and Systems Engineering, University of Seville, Spain. From January 2009 to December 2010 he was a postdoctoral fellow in the Automatic Control Laboratory, ETH Zürich, Switzerland. From March 2012 to June 2012 and from August 2013 to September 2013 he was visiting scholar in Prof. Braatz Group, Department of Chemical Engineering, MIT, USA. Since December 2010 he is Assistant Professor at University of Pavia, Italy.

He is the author or coauthor of more than 50 papers published in refereed journals, edited books, and refereed conference proceedings. His current research interests include control of nonlinear constrained systems, robust control, networked control, fault-tolerant control, autonomous surveillance and control of glycemia in diabetic patients.