Abstract—Active fault diagnosis (AFD) is crucial for safe, reliable, and high-performance operation of complex technical systems. For linear systems with uncertainties bounded within zonotopes, this paper addresses the AFD problem under closed-loop control. The AFD method reported in [17] is extended to design the reference signal of a feedback controller such that fault diagnosis is guaranteed within a prespecified time horizon. Designing the reference signal of a controller, instead of directly designing the system input, will enable application of the AFD method to systems with existing feedback controllers. The reference design problem seeks to separate the most likely model from the rest of the fault model candidates. The presented AFD method is implemented in a moving-horizon fashion through estimating the probability that each nominal/fault model can be active via sample-based Bayesian estimation. It is shown that the AFD method can enhance fault diagnosis in terms of reducing the number of time steps required for guaranteed model separation.

I. INTRODUCTION

The detection and isolation of system faults due to, for example, component failures, parameter drifts, and sensor and actuator malfunctions is key to safe, reliable, and high-performance operation of complex technical systems in a wide range of engineering applications [4]. However, system uncertainties such as disturbances and measurement noise as well as the corrective action of a feedback controller can severely impede reliable fault diagnosis.

Model-based approaches to fault diagnosis can be generally classified as active or passive methods [5], [6], [14], [22]. Passive fault diagnosis involves comparing the input-output data gathered during nominal system operation with models of potential fault scenarios, whereas active fault diagnosis (AFD) involves injecting an auxiliary input signal into the system to improve the diagnosability of potential faults with minimal impact on the nominal system operation. Several AFD methods have been reported to address the input design problem for fault diagnosis of linear systems with bounded uncertainties [3], [11], [13]. These methods generally seek to design an input that guarantees fault diagnosis within a prespecified time horizon, or conclude that no such input exists. An AFD method for linear systems with (possibly) high state dimension and/or multiple fault models is presented in [18] using the properties of the zonotopes [19]. The AFD problem has also been addressed for nonlinear systems with bounded [1] and probabilistic [9], [10], [15] uncertainties.

Closed-loop AFD methods that incorporate online system information into input design are presented in [2], [7], [15], where the input design problem is solved online at every measurement sampling time. A closed-loop AFD method is presented in [12] for designing a feedback control law that enhances the fault diagnosability. Alternatively, in [16], [21], the criterion for the online input design is defined in terms of both the nominal control objectives and the risks associated with incorrect fault diagnosis.

This paper addresses the problem of input design for fault diagnosis under closed-loop control for linear time-invariant systems subject to uncertain initial conditions, process noise, and measurement noise. The system uncertainties are considered to be bounded within constrained zonotopes [19]. The paper extends the closed-loop AFD method in [17] to design the reference signal of a feedback controller, instead of designing directly the input signal, for guaranteed fault diagnosis within a prespecified time horizon. The reference design problem seeks to separate the reachable sets of the nominal/fault models, so that system measurements only agree with predictions of a single model. Designing the reference signal of a controller will enable application of the AFD method to systems with existing feedback controllers. As in [15], [17], the online AFD problem is solved in a moving-horizon manner. However, the probability that each nominal/fault model can be active is estimated at every sampling time via sample-based Bayesian estimation. Thus, in contrast to [17], only the most likely model (i.e., the model with the highest estimated probability) is separated from the rest of the fault model candidates. This not only will lead to improved diagnosability in terms of the number of time steps needed for guaranteed model separation, but also will significantly reduce the computational complexity of the AFD problem, as shown in this paper. The effectiveness of the proposed method is demonstrated using several numerical examples.

II. PROBLEM SETUP AND PRELIMINARIES

A. Problem Formulation

Consider a discrete-time, affine system that can be described, at each time step \( k \geq 0 \), by one of the known and
observable \( n_m \) models

\[
\begin{align*}
x^{i}_{k+1} &= A^i x^i_k + B^i u^i_k + r^i_k + B^i w^i_k, \\
y^i_k &= C^i x^i_k + s^i_k + D^i w^i_k,
\end{align*}
\]

(1a)

where \( x^i_k \in \mathbb{R}^{n_x}, u^i_k \in U \subset \mathbb{R}^{n_u}, \) and \( y^i_k \in \mathbb{R}^{n_y} \) denote the system states, inputs, and outputs, respectively. \( x^i_0 \in \mathbb{X}^i \subset \mathbb{R}^{n_x} \) denotes the initial states, and \( w^i_k \in W \subset \mathbb{R}^{n_w} \) and \( v^i_k \in V \subset \mathbb{R}^{n_v} \) denote the process and measurement noise, respectively.

The constant vectors \( r^i_k \) and \( s^i_k \) describe additive faults, and are considered to be known. \( \mathbb{X}^i, U, W, \) and \( V \) are constrained zonotopes that are known a priori (see Section II-B). This work considers the class of systems where \( \mathbb{X}^i \) is assumed to be invertible for all \( i \in I = \{1, \ldots, n_m\} \). In (1), \( i = 1 \) represents the nominal system model.

The nominal system is in closed loop with a feedback controller designed to stabilize the system at the equilibrium \( (\tilde{u}^i_k, \tilde{x}^i_k) \). Hence, the inputs are defined by

\[
u^i_k = \tilde{u}^i_k - \mathbf{K} (\tilde{x}^i_k - \tilde{x}^i_k),
\]

(2)

where

\[
\tilde{x}^i_k = (C^i)^{-1} (y^i_k - s^i_k)
\]

(3)

is an estimate of the states \( x^i_k \) computed assuming that the measurement noise \( v^i_k \) is zero. Note that

\[
\tilde{x}^i_k = x^i_k + (C^i)^{-1} D^i v^i_k.
\]

(4)

The matrix \( \mathbf{K} \) in (2) is designed to control the system under nominal conditions, so that \( A^i - \mathbf{B}^i \mathbf{K} \) is Schur. The pair \( (\tilde{u}^i_k, \tilde{x}^i_k) \) denotes the equilibrium point (i.e., steady state) that the controller must reach. The relationship between the equilibrium inputs and states is given by

\[
\tilde{x}^i_k = (I - A^i)^{-1} \mathbf{B}^i \tilde{u}^i_k,
\]

(5)

where \( (I - A^i) \) is assumed to be invertible and is computed considering that the process noise is zero. Substituting (5) into (2) results in

\[
u^i_k = \tilde{u}^i_k - \mathbf{K} (\tilde{x}^i_k - (I - A^i)^{-1} \mathbf{B}^i \tilde{u}^i_k)
\]

\[
= -K \tilde{x}^i_k + (I + K (I - A^i)^{-1} \mathbf{B}^i) \tilde{u}^i_k
\]

\[
= -K \tilde{x}^i_k + \Upsilon \tilde{u}^i_k,
\]

(6)

where \( \Upsilon = (I + K (I - A^i)^{-1} \mathbf{B}^i) \). When a fault occurs, until it is diagnosed, the fault dynamics \( i \) are subject to the input

\[
u^i_k = \tilde{u}^i_k - \mathbf{K} (\tilde{x}^i_k - \tilde{x}^i_k)
\]

\[
= -K \tilde{x}^i_k + \Upsilon \tilde{u}^i_k,
\]

(7)

with \( \tilde{x}^i_k \) defined as in (3). Note that \( \tilde{u}^i_k, \mathbf{K}, \) and \( \tilde{x}^i_k \) are determined based on the nominal system dynamics. Substitution of (4) and (7) into (1a) yields

\[
x^{i}_{k+1} = A^i x^i_k + B^i u^i_k + r^i_k + B^i w^i_k,
\]

(8)

\[
y^i_k = C^i x^i_k + s^i_k + D^i w^i_k,
\]

(9)

where

\[
\mathcal{A}^i = A^i - B^i \mathbf{K},
\]

(10)

\[
\mathcal{B}^i = B^i \Upsilon,
\]

(11)

\[
\mathcal{D}^i = -\mathbf{K} (C^i)^{-1} D^i.
\]

(12)

The objective of the proposed AFD method is to determine which model in (1) describes the actual system dynamics. At a given time instant, \( i \in I \) is assumed to be the most likely model active on the system. To maximize the probability of fault diagnosis, in contrast to [17], [18], we look to determine the shortest input sequence of length \( N \) that can separate the evolution of the model \( i \) from the collection of models in (1). Since multiple input sequences of minimal length \( N \) may satisfy this requirement, we aim to select the input sequence that minimizes a given diagnosis cost. Moreover, rather than injecting the entire input sequence to the system, only the first element of the input sequence is applied to the system, and a new input sequence is recomputed at every measurement sampling time to use the newly available information. Thus, the AFD problem is implemented in a closed-loop fashion. If, at a given time instant, the model \( i \) is discarded, AFD will involve separating the most likely model among the remaining ones.

In contrast to [17], the entire input sequence \( \bar{u}^i = (u^i_0, \ldots, u^i_{N-1}) \) cannot be used as decision variables in the AFD problem, rather only the reference sequence \( \bar{u}^i = (\bar{u}^i_0, \ldots, \bar{u}^i_{N-1}) \) can be freely manipulated. This is because of the presence of the feedback component \( -\mathbf{K} \bar{x}^i_k \) in (2). In fact, in systems under closed-loop control, it is much more likely to be able to modify the reference signal of feedback controllers, instead of injecting a customized input sequence that would undermine the existing feedback controllers. In this work, it is assumed that only one model is active during the diagnosis period \( [0, \ldots, N] \). That is, the fault diagnosis is fast enough so that the switching between models can be neglected during diagnosis.

B. Constrained Zonotopes and Set Operations

The formulation of the proposed AFD problem relies on a new class of sets introduced in [19] as an extension of the zonotopes: constrained zonotopes.

Definition 1: A set \( Z \subset \mathbb{R}^n \) is a constrained zonotope if there exists \( (G, c, A, b) \in \mathbb{R}^{n \times n_z} \times \mathbb{R}^n \times \mathbb{R}^{n_z \times n_x} \times \mathbb{R}_+^n \) such that \( Z = \{G \xi + c : \|\xi\|_{\infty} \leq 1, A \xi = b\} \). The columns of \( G \) are called the generators, \( c \) is the center, and \( A \xi = b \) are the constraints.

In contrast to standard zonotopes, Definition 1 allows linear equality constraints on \( \xi \). Hence, it can be used for representing any convex polytope [19]. We use the shorthand \( Z = \{G, c, A, b\} \) and \( Z = \{G, c\} \) for constrained and standard zonotopes, respectively, which are called CG-representation of the sets. The CG-representation has two primary advantages. First, it simplifies the computation of some important set operations. For \( Z = \{G_x, c_x, A_x, b_x\}, W = \{G_w, c_w, A_w, b_w\} \subset \mathbb{R}^n, Y = \{G_y, c_y, A_y, b_y\} \subset \mathbb{R}^k, \) and \( R \in \)
Constrained zonotopes are closed under linear mappings (13), Minkowski sums (14), and intersections (15) as well as under cartesian products. The second advantage of the CG-representation is that there exists efficient methods for enclosing a given constrained zonotope within another constrained zonotope of lower complexity (i.e., with fewer generators and constraints) [19]. This is a useful feature since the results of set operations (13)–(15) can become more complex than their arguments.

C. Reachable Set Notation and Computations

A tilde denotes a sequence associated with (8)–(9), for example, \( \tilde{u}^i \equiv (\tilde{u}_0^i, \ldots, \tilde{u}_{N-1}^i) \). The notation \( \tilde{u}^i_{0:N-1} \) refers to the time horizon \( 0 : N - 1 \). We introduce the shorthand \( \lambda^i \equiv (x_0^i, \tilde{w}, \tilde{v}) \), where \( \lambda^i \) is a random variable distributed according to an arbitrary probability distribution in the set \( \Delta^i = X_0^i \times \tilde{W} \times \tilde{V} \), where \( \tilde{W} = W \times \cdots \times W \) denotes the cartesian product of sets \( W \). Note that \( \lambda^i \) contains all random variables that affect the output at time \( N, Y_N^i, \forall i \in I \). For each model \( i \in I \) and time \( k \geq 0 \), define the state solution map \( \Phi_k^i: R^{kn \times n} \times R^{n \times n \times n} \rightarrow R^{kn} \) and the output solution map \( \psi_k^i: R^{kn \times n} \times R^{n \times n \times n} \rightarrow R^{ny} \), so that \( \Theta_k^i(\tilde{u}^i, \lambda^i) \) is the state of (8) and \( \Psi_k^i(\tilde{u}^i, \lambda^i) \) is the output of (9) at \( k \) given the initial state \( x_0^i \), reference input sequence \( \tilde{u}^i \), process noise \( \tilde{w} \), and measurement noise \( \tilde{v} \). Define \( \Theta_k^i \equiv (\Theta_0^i, \ldots, \Theta_k^i) \) and \( \Psi_k^i \equiv (\Psi_0^i, \ldots, \Psi_k^i) \). The notation \( \Theta_0^i \) and \( \Psi_0^i \) is used to refer to the time horizon \( 0 : k \).

Definition 2: The state reachable and output reachable set for model \( i \) at time \( k \) are defined, respectively, by

\[
\Theta_k^i(\tilde{u}^i, \lambda^i), \Delta^i \equiv \{ \Phi_k^i(\tilde{u}^i, \lambda^i) : \lambda^i \in \Delta^i \},
\]

\[
\Psi_k^i(\tilde{u}^i, \lambda^i), \Delta^i \equiv \{ \Psi_k^i(\tilde{u}^i, \lambda^i) : \lambda^i \in \Delta^i \}.
\]

For brevity, explicit dependence of the reachable sets on \( \Delta^i \) will be omitted. Iterating (8)–(9) \( k \) times yields matrices \( \Theta^i, \Theta^i_0, \Theta^i_k \) and \( \Psi^i, \Psi^i_0, \Psi^i_k \) such that

\[
\Phi_k^i(\tilde{u}^i, \lambda^i) = \Theta_{k}^i x_0^i + \Theta_{k}^i \tilde{u}^i + \Theta_{k}^i \tilde{w} + \Theta_{k}^i \tilde{v}.
\]

It follows that the reachable sets can be expressed as

\[
\Phi_k^i(\tilde{u}^i) = \Theta_{k}^i x_0^i + \Theta_{k}^i \tilde{u}^i + \tilde{r} + \tilde{b}_w \tilde{w} + \tilde{b}_v \tilde{v},
\]

\[
\Psi_k^i(\tilde{u}^i) = \tilde{C} \tilde{d}_k^i(\tilde{u}^i) + \tilde{d}_v^i \tilde{v}.
\]

Since \( X_0^i \), \( W \), and \( V \) are constrained zonotopes, these equations imply that \( \Phi_k^i(\tilde{u}^i) \) and \( \Psi_k^i(\tilde{u}^i) \) are constrained zonotopes and, therefore, can be computed efficiently using (13)–(14) (see [17] for more details).

III. OPEN-LOOP INPUT DESIGN

This section presents the open-loop reference design problem. Consider model \( \tilde{i} \in I \) to be the most likely model (i.e., active model) of the system. The AFD objective is to determine a reference input sequence that guarantees separation of the reachable set of \( \tilde{i} \) from the reachable set of all the other models. For a given horizon \( N \), this requires verifying

\[
\Psi_{0:N}^i(\tilde{u}^i, \Delta^i) \cap \Psi_{0:N}^{\tilde{i} \neq i}(\tilde{u}^i, \Delta^i) = \emptyset
\]

for all \( j \in I \) with \( j \neq \tilde{i} \). When model \( j \) is active on \([0, N]\), then the pair \( (\tilde{u}^i, \tilde{y}) \) is guaranteed to satisfy

\[
\tilde{y} \in \Psi_{0:N}^{j \neq i}(\tilde{u}^i, \Delta^i).
\]

Conversely, if (21) does not hold, the model \( j \) would not be active on \([0, N]\). The set of all inputs satisfying (20) is denoted by \( \mathcal{J}_N(\tilde{i}, X_0^i) \). For the input set \( \mathcal{J}_N(\tilde{i}, X_0^i) \), we define the input design problem in terms of the length of the input sequence and its norm. Let \( N \) be the smallest value for which \( \mathcal{J}_N(\tilde{i}, X_0^i) \) is non-empty. Assuming \( N < \infty \), an optimal input separating \((\tilde{i}, X_0^i)\) in \( N \) steps is defined as the solution of

\[
\inf \{ J_N(\tilde{u}^i) : \tilde{u}^i \in U_N \cap \mathcal{J}_N(\tilde{i}, X_0^i) \},
\]

where \( \tilde{u}^i = (u_0^i, \ldots, u_{N-1}^i) \) and \( U_N \equiv U \times \cdots \times U \). According to (4) and (7),

\[
u_k^i = -K(x_k^i + (C^i)^{-1}D_v^i v_k^i) + \bar{Y}^i_k.
\]

Since \( \tilde{u}^i \) is the only degree of freedom in the input design problem, the satisfaction of the input constraints (i.e., \( u_k^i \in U \)) can be enforced only in terms of \( \tilde{u}^i_k \)

\[
-K(x_k^i + (C^i)^{-1}D_v^i v_k^i) + \bar{Y}^i_k \in U,
\]

\( \forall \bar{x}_k^i \in \Phi_k^i(\tilde{u}^i_{0:k}) \) and \( \bar{v}_k^i \in V \). According to the definition of the reachable sets, (24) can be rewritten as

\[
-K(\Phi_k^i(u_{k-1}^i)) + (C^i)^{-1}D_v^i v_k^i) + \bar{Y}^i_k \in U,
\]

(25) which, based on (18), takes the form

\[
-K(\tilde{u}^i_{0:k-1}) - K(\Phi_k^i(0)) + (C^i)^{-1}D_v^i V) + \bar{Y}^i_k \in U.
\]

Isolating the terms containing \( \tilde{u}^i \) leads to

\[
-K(\tilde{u}^i_{0:k-1}) - Y^i_k \in U \setminus K \left[ \Phi_k^i(0) (C^i)^{-1}D_v^i V \right],
\]

(27) where the symbol \( \setminus \) denotes the Pontryagin difference. Thus, problem (22) must be solved subject to the input reference constraints (27). We define the cost function as

\[
J_N(\tilde{u}^i) = \sum_{k=1}^n \tilde{u}_{k}^i \tilde{r}_{k-1}^i
\]

where \( R \in R^{n \times n} \) is a positive semidefinite matrix. Note that when the set (27) becomes too complex, the order of the set can be reduced via outer approximations of constrained zonotopes.

As in [17], the optimization problem can be formulated as a mixed integer quadratic program (MIQP), which can be solved using, for example, CPLEX. Note that the minimum
separation horizon $N$ can be computed by repeating the above optimization procedure with $N$ increasing from 1 until either a feasible program is generated or a prespecified $N$ is reached. Since the number of binary variables is proportional to the total number of generators in the constrained zonotopes that define $S_N(i, X_i)$, constrained zonotopic reduction techniques can be used to reduce the computational complexity of the optimization problem (e.g., see [19]). The improved computational efficiency in this case comes at the expense of increased conservatism of the designed input. However, it is shown in [17] that large computational gains can be achieved with only a slight increase in conservatism.

In comparison with the AFD problem presented in [17], the main differences of the above formulation arise from: (i) designing the reference signal of a feedback controller, rather than the control inputs, in order to enable application of the AFD method to systems with existing feedback controllers, and (ii) separating the most likely reachable set only. In particular, the second feature allows for reducing the complexity of the input design problem, as the problem size grows linearly with the number of models and no longer combinatorially as in [17]. Note that when the model $i$ is discarded, a new optimization based on the most likely model among the remaining ones must be solved for active fault diagnosis.

IV. CLOSED-LOOP INPUT DESIGN

This section considers the AFD problem in a moving-horizon framework under the assumption that $\hat{u}_{k+1}^{[1]}$ can be designed during the time interval $[k, k+1]$ with full knowledge of the measurements $\bar{y}_{0,k}$. In this case, the information provided by the measurements $\bar{y}_{0,k}$ at each $0 \leq k < N$ is used to improve the effectiveness of the designed reference inputs. In particular, a bank of set-valued observers is used to reduce the uncertainty on the state reachable sets. Set-valued observers are observers whose states at $k+1$ are sets $\hat{X}_{k+1|k}^i \subset \mathbb{R}^n$, which are guaranteed to contain all $X_{k+1|k}$ consistent with (8)-(9), the constraints $x_{0}^{i} \in X_{0}^{0}$ and $(w_{k}, v_{k}) \in W \times V$, and the measured output values $y_{k}$ for all $k \leq K$ [17], [20]. Each observer uses all information available at time $k$ to update the predicted state $\hat{X}_{k+1|k}$ using the measurement $y_{k+1}$, yielding $\hat{X}_{k+1|k+1}^i$. For each model $i \in I$, the set-valued observer is defined recursively as

$$\hat{X}_{k+1|k}^i \supset \hat{X}_{k|k-1}^i \cap \left\{ x^i : C_i x^i \in y_k - s^{i} \ominus (-D_i) V \right\}, \quad (28)$$

$$\hat{X}_{k+1|k}^i \supset \bigoplus_{i=1}^{m} \hat{X}_{k|k-1}^i \ominus \bigoplus_{i=1}^{m} \bar{u}_{k}^{[i]} \ominus r^{i} \ominus B_{i}^{[i]} W \ominus \bigoplus_{i=1}^{m} \bar{v}_{i}^{[i]} V, \quad (29)$$

with $\hat{X}_{0|0}^{i} \equiv X_{0}^{i}$ and $\ominus$ representing the Minkoswki sum. It is assumed that the constrained zonotopic enclosures $X_{0}^{i}$ are available prior to $k = 0$. In principle, the constrained zonotope observer is exact. In practice, however, it is necessary to periodically apply conservative reduction techniques. Thus, for each $k$, $\hat{X}_{k+1|k}$ is reduced to a target number of generators and constraints using the methods presented in [19].

A. Moving-Horizon Implementation

The input reference can be designed in closed loop by solving the open-loop AFD problem recursively in a moving-horizon fashion. At time $k = 0$, the set-valued observers (28)-(29) are initialized via $\hat{X}_{0|k-1}^{i} = X_{0}^{0}$, $\forall j \in I$. The model $\hat{i}$, which is the most likely model to be active on the system, is chosen via Bayesian estimation, as described in Section IV-B. Then, a minimum separation horizon $N$ and an optimal input reference $\hat{u}_{k|N-1}$, separating $\hat{i}$ from all the other models $j \in I$ in $N$ steps, is computed using the procedure of Section III. According to the receding-horizon paradigm, only $\hat{u}_{0}^{[i]}$ is applied to the system and $\hat{X}_{0|1}$ is computed. At time $k = 1$, the set-valued observer $\hat{X}_{k|1}^{i}$ is updated using the measurement $y_{1}$ for each $j \in I$ (see (28)-(29)). Now, three scenarios can occur: 1) $y_{1}$ is consistent with one model only. In this case, the fault is isolated and the procedure ends; 2) $y_{1}$ is not consistent with $\hat{i}$, but with at least two models $j \in I$. In this case, $\hat{I}$ is updated to contain the non-discarded models only. A new most likely model is chosen according to the selection criterion (see Section IV-B) among the remaining ones and a new optimal input reference is computed; and 3) $y_{1}$ is consistent with $\hat{i}$ and at least another model $j \in I$. In this case, $\hat{i}$ remains to be the most likely model. As in the previous case, $\hat{I}$ contains the non-discarded models only. A new optimal input reference sequence separating $\hat{i}$ from the other models is computed at $k = 1$. Note that due to the outer approximations used in the set-valued observers, there is no guarantee that the input sequence computed at this time would be separating the reachable sets of the competing models in $N - 1$ steps. In this case, one can always apply the remaining of the reference input sequence computed at $k = 0$. This procedure is repeated for $k > 0$ until a guaranteed diagnosis is provided. It can be shown that this approach leads to fault diagnosis in a finite number of steps under mild assumptions [8].

B. Sample-based Bayesian Estimation

The input reference is designed based on the knowledge of the most likely model active on the system (see Section III). To this end, a sample-based Bayesian estimator is designed to determine the most likely model. The estimator consists of two steps. The intersection in (28) is first computed at each time step $k$. Given the measurements $y_{k}$ at time $k$, the probability of a certain model $i$ being active on the system at time $k$ is then estimated via

$$P(i_{k} | y_{k}) = \frac{f(y_{k} | i_{k}) P(i_{k} | \bar{y}_{0,k-1})}{\sum_{i=1}^{m} f(y_{k} | i_{k}) P(i_{k} | \bar{y}_{0,k-1})},$$

where $P(i_{k} | \bar{y}_{0,k-1})$ is initialized as $P(i_{0}) = 1/m$, and $f(y_{k} | i_{k}) = \int_{\xi} f(y_{k} | i_{k}, \xi) f(\xi | i_{k}) d\xi$. The latter probability distribution is approximated via sampling $C_{k}^{i} \hat{X}_{k|0}^{i}$ according to the distribution $\Delta^{i}$. The measurements $y_{k}$ are then used to determine the samples lying in the intersection of

$$\hat{X}_{k+1|k}^{i} \cap \left\{ x_{k}^{i} : C_{k}^{i} x_{k}^{i} \in y_{k} - s^{i} \ominus (-D_{k}^{i}) V \right\},$$
where the probability of each of these samples can be computed using the distribution defined on $V$. Ultimately, the probability distribution $f(y_i|k)$ is approximated as the sum of the weighted samples.

V. NUMERICAL ILLUSTRATIONS

The performance of the proposed AFD method is demonstrated using three numerical examples.

A. Example 1

In this example, the open-loop computational cost of the proposed AFD method is compared to that in [17]. Note that the proposed method aims to separate the most probable model from the rest of the models. Here, we consider $K = 0$. In this case, the AFD involves designing the system input sequence, rather than the reference input. We consider a 10-state system and three possible models, namely $A_1$, $A_2$, and $A_3$.

$$A_1 = \begin{bmatrix}
0.5 & 0.5 & -0.5 & -0.5 & 0 & 0 & 1 & 1 & -0.1 & 0.1 \\
0.5 & 0.5 & -0.5 & -0.5 & 1 & 1 & 0 & 0 & -0.1 & 0.1 \\
-0.5 & 0.5 & -0.5 & -0.5 & 0 & 0 & 1 & 0 & -0.1 & 0.1 \\
-0.5 & -0.5 & -0.5 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0.1 \\
-0.5 & 0.5 & -0.5 & -0.5 & 1 & 1 & 1 & 1 & 0.1 & -0.1 \\
-0.5 & -0.5 & -0.5 & -0.5 & 0 & 0 & 0 & 0 & 0.1 & -0.1
\end{bmatrix}$$

$A_2 = A_1$ with $A_2^{(1.5)} = 0$, $A_3 = A_1$ with $A_3^{(8.10.9.10)} = 0$, where $A_1^{(8.10.9.10)}$ indicates the $\mathbb{R}^{3 \times 2}$ submatrix of $A$. For all $i \in \{1, 2, 3\}$, $B_i = 0.5I$, $B_i^w = 0.05I$, and $C_i = D_i = I$. $I_{10}$ denotes a column vector with 10 elements all equal to 1 and $I$ denotes the 10-dimension identity matrix. The initial condition, process noise, and measurement noise sets are, respectively, given by

$$V = \{0.1I, 0.1_{10}\},$$

$$X_0^{[i]} = \{0.4[1_{10}], 0 \cdot 1_{10}, 0.1_{10}, 1]^{T}, 1\},$$

$$W = \{0.6[1_{10}], 0 \cdot 1_{10}, 0.5 \cdot 1_{11}, -1\}.$$

Table I shows the comparison between the average computation time required to solve the open-loop optimization problem in [17] and the proposed AFD problem. As can be seen, the computation time of the proposed method is approximately one third of that of the method in [17]. When

<table>
<thead>
<tr>
<th></th>
<th>Proposed AFD Method</th>
<th>AFD Method in [17]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = 0$</td>
<td>2.23 s</td>
<td>6.49 s</td>
</tr>
<tr>
<td>$K \neq 0$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE II


<table>
<thead>
<tr>
<th></th>
<th>Proposed AFD Method</th>
<th>AFD Method in [17]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 4.1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. Example 2

In this example, we compare the input sequence designed using the AFD method described in Section IV with that of [17]. One of the key features of the proposed closed-loop AFD method is the use of a sample-based Bayesian estimation approach to determine the most likely model at each time step. Consider five models defined by

$$A_{1,4,5} = [0.6, 0.2], A_2 = [0.6, 0.7], A_3 = [0.6, 0.2],$$

$$B_{1,2,3,5} = [-0.3861, 0.1994], B_4 = [-0.3861, 0.1994],$$

$$C_{1,2,3,4} = [1, 0], C_5 = [1, 0],$$

and, for all $i \in \{1, 2, 3, 4, 5\},$

$$B_w^i = [0.1215, 0.0598, 0.1215], D_y^i = [1_0], r^i = s^i = [0_0], K = 0.$$

The measurement noise is considered to lie in a zonotope $V = \{0.2 \cdot I, 0_0\}$. For all $i \in I$, $W = \{0.5, 0.5, 0, 0.5, 0\}$. Tables II and III compare the performance of the proposed closed-loop AFD method to that of [17] in terms of the length of the input sequence and its norm. As can be seen, the proposed method allows for reducing both the length and norm of the input sequence required for isolating the fault. This is achieved while the computation time is significantly lower than that of the closed-loop AFD method in [17].

<table>
<thead>
<tr>
<th></th>
<th>Proposed AFD Method</th>
<th>AFD Method in [17]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = 0$</td>
<td>4.47</td>
<td>4.86</td>
</tr>
<tr>
<td>$K \neq 0$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE III

PERFORMANCE COMPARISON OF THE PROPOSED CLOSED-LOOP AFD METHOD AND THAT IN [17] FOR THE CASES OF $K = 0$ AND $K \neq 0$ IN EXAMPLE 3.
C. Example 3

In this example, we compare the effectiveness of closed-loop AFD when $K = 0$ and when a state feedback compensation, which may mask the faults.

Three models are considered

$$A_i^1 = \begin{bmatrix} -0.6 & 0.2 \\ -0.2 & 0.7 \end{bmatrix}, \quad A_i^2 = \begin{bmatrix} 0.6 & 0 \\ -0.2 & 0.7 \end{bmatrix}, \quad A_i^3 = \begin{bmatrix} 0.6 & 0.2 \\ 0 & 0.7 \end{bmatrix},$$

and

$$B_i^1 = \begin{bmatrix} -0.3861 & 0.1994 \\ -0.1994 & 0.3861 \end{bmatrix}, \quad C_i^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B_i^2 = \begin{bmatrix} 0.1215 \\ 0.0598 \end{bmatrix}, \quad 0.1215,$$

$$D_i^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad r_i^1 = s_i^1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

for all $i \in \{1, 2, 3\}$. The feedback controller is defined by

$$K = \begin{bmatrix} -0.9982 & -0.1638 \\ -0.4189 & 1.0323 \end{bmatrix}.$$  

The measurement noise is considered to lie in a zonotope $V = \{0.2 \cdot I, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\}$. For all $i \in I$,

$$W = \left\{ \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0.5 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.5 & 0.5 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.5 & 0.5 \\ 0 & 0 \end{bmatrix} \right\},$$

$$X_0^i = \left\{ \begin{bmatrix} 0.1 & 0.1 \\ 0 & 0.1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.5 & 0.5 \\ 0 & 0 \end{bmatrix} \right\}.$$

Table IV shows the comparison between the two cases in terms of the length of the designed input sequence for fault diagnosis. It is interesting to notice that the case of $K \neq 0$ takes longer to isolate the fault. While in industrial processes it is much more likely to be able to design a reference input rather the entire input sequence, this may come at the expense of a more difficult isolation due to the feedback compensation, which may mask the faults.

VI. CONCLUSIONS

A probabilistic framework is presented for active fault diagnosis under closed loop. The active fault diagnosis problem involves designing the reference signal of a feedback controller to achieve guaranteed fault diagnosis within a pre-specified time horizon. The reference input design problem is implemented in a receding-horizon fashion. Simulation results indicate that online design of the reference signal not only will enhance the effectiveness of active fault diagnosis, but also will significantly improve the computational efficiency of the active fault diagnosis problem.

ACKNOWLEDGMENT

The authors thank Prof. G. De Nicolao for his technical comments.