2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

Semidefinite Programming Relaxation of Optimum Active Input Design for Fault Detection and Diagnosis: Model-based Finite Horizon Prediction

Kwang-Ki K. Kim and Richard D. Braatz, *IEEE Fellow*

Abstract— **This paper establishes optimal/suboptimal *active* fault detection and diagnosis (FDD) methods in which semidefinite programming relaxation is used and the optimality criteria are information theoretic measures of the statistical distance between probability distributions. The design problems are formulated as optimizations in which an optimal sequence of inputs within a prediction horizon is computed for maximizing the statistical discrimination of different models of fault scenarios. Three different measures for the degree of statistical distinguishability between two hypothesized stochastic dynamical system models are considered and their mathematical properties that are related to Bayesian hypothesis tests are studied. The resulting input design problems are non-convex and we propose associated convex relaxation methods that can be solved in polynomial time using interior point methods. Numerical simulations with an aircraft model are provided to illustrate and demonstrate the presented methods of optimal input design for FDD.**

I. INTRODUCTION

The complexity of devices and processes implies that faults are inevitable, and the tight interactions between instrumentation and other components of the overall system can result in cascading effects with significant economic, environmental, and human damages. To properly and safely operate the facilities and devices in real-time while preventing any unallowable behaviors of the system, reliable FDD algorithms are needed that monitor the inputs and outputs of the system and determines whether a fault occurs and to point to the location of the fault (aka *fault diagnosis*). In addition, without an optimal integration between the monitoring and control systems, the response to faults can reduce reliability and profit or can be overly conservative, for example, by initiating an unnecessary automated shutdown of the facility due to false alarm.

The design of FDD procedures are challenged by the presence of disturbances, noise, and model uncertainties that can make the symptoms of faults/failures indiscernible. Classical *passive* FDD methods monitor the observables of the system and make a decision on whether and where a fault has occurred, whereas *active* FDD approaches intentionally intervene in the operations and attempt to excite or perturb the observables such that abnormal behaviors are exhibited [1]–[5]. There have been many research efforts to suggest systematic methods for such auxiliary input design, both in the stochastic system setting [6] and in the deterministic uncertain system setting [7]–[9]. The effects of feedback in terms of the performance of FDD and quadratic cost optimality criteria have been investigated [10], [11], as have been finite- or infinite-horizon control methods for the design of active input signals for FDD [5], [6], [9].

This paper considers three different measures for the difference between two hypothesized dynamical system models.* The active input design problems are formulated as optimizations for which an optimal input or a sequence of inputs is computed to maximize distinguishability (or discrimination) of different models of fault scenarios. To quantify discrimination of two stochastic dynamical system models, we use information theoretic measures that compute statistical distances between the solutions of the associated stochastic differential equations. The resultant optimizations are non-convex (more precisely, concave programs) that are NP-hard. We propose semidefinite programming relaxations for those non-convex optimizations, which provide suboptimal solutions for active input design problems with guaranteed bounds of performance degradation for certain special cases (e.g., when there are only box constraints on the inputs).

An underlying assumption considered in this paper is that the procedures of FDD are based on a statistical decision that solves Bayesian inference problems for which the measurements of observables are used to infer a hidden process. In this paper, robustness is considered with respect to stochastic uncertainties, disturbances, and noises. Robust FDD can be considered as maximizing the confidence of a binary decision (for fault detection) and locating a correct hypothesis (for fault diagnosis) among many candidate fault scenarios in the presence of uncertainty in a given data set. The objective of active input design for FDD is to facilitate the associated statistical decision and maximize its robustness.

Notation: The following notation is used throughout this paper: $\mathbf{Pr}$ denotes probability; $\mathbf{E}$ or $\overline{(\cdot)}$ denotes mathematical expectation or mean; $\mathcal{N}(a,b)$ is the Gaussian distribution with the mean a and the covariance b; the symbol $\sim$ means "distributed as". $\mathbb{S}^n \subset \mathbb{R}^{n\times n}$ refers to the set of real symmetric matrices and its subsets $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ are used to denote the set of positive semidefinite and definite matrices, respectively. Equivalent notations are $X \succeq 0$ ($X \succ 0$) for $X \in \mathbb{S}^n_+$ ($X \in \mathbb{S}^n_{++}$).

II. PRELIMINARIES

A. Bayesian Hypothesis Testing

This section provides a concise introduction to Bayesian hypothesis testing. For details of Bayesian theory, readers are referred to [12], [13], for example. Consider the set of hypotheses $\mathcal{H} \triangleq \{H_0, H_1, \cdots, H_m\}$ in which the i^{th} hypothesis H_i corresponds to a model to explain the observed data z. A Bayesian hypothesis testing problem is to find an optimal hypothesis $H^\star$ that is most consistent with the observed data z in the sense that it maximizes the associated

K.-K. K. Kim is with the University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA, kkim78@illinois.edu, and Massachusetts Institute of Technology, Cambridge, MA, 02139, USA, kwangki@mit.edu

R. D. Braatz is with the Massachusetts Institute of Technology, Room 66-372, Cambridge, MA, 02139, USA, braatz@mit.edu.

---

[*] A measure for difference between two hypothesized dynamical system models is to quantify the difference between two probability distribution functions associated with random processes that are the solutions of stochastic differential (or difference) equations.

978-3-952-41734-8/©2013 EUCA

posterior distribution $p_{\mathbf{H}|\mathbf{z}}(h(z)|z)$ where $h : \mathcal{Z} \to \mathcal{H}$ is a deterministic decision rule. This problem can be formulated as an optimization

$$H^{\star}(z) := \arg \max_{H_i \in \mathcal{H}} p_{\mathbf{H}|\mathbf{z}}(H_i|z). \tag{1}$$

For a more general problem formulation, a similar optimization can be written as

$$\begin{aligned} &\min_{h(\cdot)\in\mathcal{H}} \mathbf{E}_{\mathbf{H}|\mathbf{z}}[C(H, f(z))] \\ &\qquad = \min_{H_i \in \mathcal{H}} \left\{ \sum_{j=0}^{m} C(H_j, H_i) p_{\mathbf{H}|\mathbf{z}}(H_j|z) \right\} \\ &\qquad = c \min_{H_i \in \mathcal{H}} \left\{ \sum_{j=0}^{m} C(H_j, H_i) p_{\mathbf{z}|\mathbf{H}}(z|H_j) p_{\mathbf{H}}(H_j) \right\} \end{aligned} \tag{2}$$

and

$$H^{\star}(z) = \arg \min_{H_i \in \mathcal{H}} \left\{ \sum_{j=0}^{m} C(H_j, H_i) p_{\mathbf{z}|\mathbf{H}}(z|H_j) p_{\mathbf{H}}(H_j) \right\} \tag{3}$$

where $C : \mathcal{H} \times \mathcal{H} \to \mathbb{R}_{+}$, $C(H, h(\cdot))$ refers to the cost of deciding that the hypothesis is $h(\cdot)$ when the correct hypothesis is H, $c = \sum_{k=0}^{m} p_{\mathbf{z}|\mathbf{H}}(z|H_k) p_{\mathbf{H}}(H_k)$ denotes the normalization factor that is independent of the hypothesis, and $p_{\mathbf{z}|\mathbf{H}}(z|H_j)$ and $p_{\mathbf{H}}(H_j)$ are the likelihood function and prior distribution associated with the hypothesis H_j, respectively.

III. PROBLEM STATEMENT

Consider the discrete-time linear time-varying stochastic system

$$\begin{aligned} x_{t+1} &= A_t x_t + B_t u_t + E_t w_t \\ y_t &= C_t x_t + D_t u_t + F_t v_t \end{aligned} \tag{4}$$

where w and v are independent Wiener processes, i.e., $w_t \sim \mathcal{N}(\mu_w, \Sigma_w)$, $v_t \sim \mathcal{N}(\mu_v, \Sigma_v)$ for all t, $\mathbf{E}[w_t w_s^{\mathrm{T}}] = 0$ and $\mathbf{E}[v_t v_s^{\mathrm{T}}] = 0$ for all $t \neq s$, and $\mathbf{E}[w_t v_s^{\mathrm{T}}] = 0$ for all t and s.

A. Bayesian Inference for FDD using Multiple Models of Fault Scenarios

Suppose that there are m fault scenarios and the model associated with the i^{th} fault scenario is given by

$$M_i : \begin{cases} x_{t+1}^i = A_t^i x_t^i + B_t^i u_t^i + E_t^i w_t^i \\ \quad y_t^i = C_t^i x_t^i + D_t^i u_t^i + F_t^i v_t^i \end{cases} \tag{5}$$

where the superscript i refers to the occurrence of the i^{th} hypothesized fault. Roughly speaking, a procedure of fault detection and diagnosis is to find the most probable model from the set of hypothesized models $\mathcal{M} \triangleq \{M_0, \cdots, M_m\}$ and Bayes' rule is applied to quantify probabilistic confidence levels as functions of the measurements for all the hypothesized models.

IV. STATISTICAL DISTANCE MEASURES FOR HYPOTHESIS TESTING

A. Distance Measure between Gaussian Hypotheses

A measure of distance between two Gaussian hypotheses can be used to characterize the performance limitation of the decision process based on the Bayesian approach.

1) Symmetrized Relative Entropy as a Distance Measure:

A common measure for the statistical distance between two probability distributions is the relative entropy, which is also called the Kullback-Leibler distance (or divergence).

Definition 1 (*See also [14] for details*). For two probability density functions f and g, relative entropy is defined by

$$d_{\mathrm{KL}}(f||g) \triangleq \int_{\mathbb{R}^n} f(x) \ln \frac{f(x)}{g(x)} dx. \tag{6}$$

For two probability mass functions f and g, relative entropy is defined by

$$d_{\mathrm{KL}}(f||g) \triangleq \sum_{x \in \mathcal{X}} f(x) \ln \frac{f(x)}{g(x)}, \tag{7}$$

where the support set $\mathcal{X}$ is assumed to be countable. For a probability mass function f and a probability density function g, relative entropy is defined by (7). For a probability density function f and a probability mass function g, relative entropy is defined by (6) and its value is indeed infinity, since the integrand is finite only if the support of f is contained in the support of g.[†]

This measure of distance between two probability distribution is not symmetric, i.e., $d_{\mathrm{KL}}(f||g) \neq d_{\mathrm{KL}}(g||f)$, in general. For example, consider two different Gaussian distribution functions $f : \mathbb{R}^n \to [0, 1]$ and $g : \mathbb{R}^n \to [0, 1]$. In particular, $f = \mathcal{N}(\mu_1, \Sigma_1)$ and $g = \mathcal{N}(\mu_2, \Sigma_2)$. Then the KL distances are

$$\begin{aligned} d_{\mathrm{KL}}(f||g) = \tfrac{1}{2} \big(&\ln\det\Sigma_2 - \ln\det\Sigma_1 + \mathrm{Tr}\,\Sigma_2^{-1}\Sigma_1 \\ &+ (\mu_1 - \mu_2)^{\mathrm{T}} \Sigma_2^{-1} (\mu_1 - \mu_2) - n \big) \end{aligned} \tag{8}$$

and

$$\begin{aligned} d_{\mathrm{KL}}(g||f) = \tfrac{1}{2} \big(&\ln\det\Sigma_1 - \ln\det\Sigma_2 + \mathrm{Tr}\,\Sigma_1^{-1}\Sigma_2 \\ &+ (\mu_1 - \mu_2)^{\mathrm{T}} \Sigma_1^{-1} (\mu_1 - \mu_2) - n \big). \end{aligned} \tag{9}$$

They are the same if $\Sigma_1 = \Sigma_2$, but not the same in general.

For a symmetric distance measure, consider one of the followings:

$$\begin{aligned} \rho_{\mathrm{KL}}^{\min}(f, g) &\triangleq \min\{d_{\mathrm{KL}}(f||g), d_{\mathrm{KL}}(g||f)\}, \\ \rho_{\mathrm{KL}}^{\max}(f, g) &\triangleq \max\{d_{\mathrm{KL}}(f||g), d_{\mathrm{KL}}(g||f)\}, \\ \rho_{\mathrm{KL}}^{\mathrm{ave}}(f, g) &\triangleq \frac{1}{2}\left(d_{\mathrm{KL}}(f||g) + d_{\mathrm{KL}}(g||f)\right). \end{aligned} \tag{10}$$

We now show the relations between the KL-divergence and the likelihood ratio test. For notational convenience, rewrite the likelihood function corresponding to the k^{th} model of the hypothesis H_k as

$$\mathcal{L}_k(z) = p_{\mathbf{z}|\mathbf{H}}(z|H_k)$$

where z refers to the concatenation of all observables. Define the ratio of two likelihood functions by

$$R_{i,j}(z) \triangleq \frac{\mathcal{L}_i(z)}{\mathcal{L}_j(z)}$$

[†] Definitions for relative entropy of two probability distributions of different types of supports, i.e., continuous and discrete support sets, were not studied in classical information theory. However, relative entropy is defined with a measure function $f(\cdot)$ for both of the continuous support case (6) and the discrete support case (7) so that it is natural to follow the support of f for definition, provided that the other probability distribution $g(\cdot)$ is well-defined over that support.

and define the (probability) measure-dependent quantity

$$T_{i,j}(\mu') \triangleq \int_{\mathcal{Z}} \ln(\max\{R_{i,j}(z), R_{j,i}(z)\}) \mathrm{d}\mu(z) \tag{11}$$

where μ' denotes the first-order derivative of the measure μ, provided it is differentiable. Note that since $\max\{R_{i,j}(z), R_{j,i}(z)\} \geq 1$ for all $z \in \mathcal{Z}$, $T_{i,j}(\mu') \geq 0$ for any measure satisfying $\mu'(z) \geq 0$ for all $z \in \mathcal{Z}$. It is straightforward to show that

$$\begin{aligned} T_{i,j}(p_{\mathbf{z}|\mathbf{H}}(z|H_i)) &= d(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j)), \\ T_{i,j}(p_{\mathbf{z}|\mathbf{H}}(z|H_j)) &= d(p_{\mathbf{z}|\mathbf{H}}(z|H_j), p_{\mathbf{z}|\mathbf{H}}(z|H_i)), \end{aligned} \tag{12}$$

which also implies

$$\begin{aligned} &\rho_{\mathrm{KL}}^{\min}(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j)) \\ &\quad = \min\left\{T_{i,j}(p_{\mathbf{z}|\mathbf{H}}(z|H_i)), T_{i,j}(p_{\mathbf{z}|\mathbf{H}}(z|H_j))\right\}. \end{aligned} \tag{13}$$

Therefore, the symmetric measure using the KL divergence $\rho_{\mathrm{KL}}^{\min}$ can be interpreted as the expectation of the logarithm of the likelihood ratio with respect to the likelihood function corresponding to the minimum value. Similarly, we have the following relations:

$$\begin{aligned} &\rho_{\mathrm{KL}}^{\max}(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j)) \\ &\quad = \max\left\{T_{i,j}(p_{\mathbf{z}|\mathbf{H}}(z|H_i)), T_{i,j}(p_{\mathbf{z}|\mathbf{H}}(z|H_j))\right\} \end{aligned} \tag{14}$$

and

$$\begin{aligned} &\rho_{\mathrm{KL}}^{\mathrm{ave}}(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j)) \\ &\quad = \frac{1}{2}\left(T_{i,j}(p_{\mathbf{z}|\mathbf{H}}(z|H_i)) + T_{i,j}(p_{\mathbf{z}|\mathbf{H}}(z|H_j))\right). \end{aligned} \tag{15}$$

2) Bhattacharyya Bound as an Upper Bound on the Bayes Risk: A Bayesian hypothesis test has a finite probability of selecting the incorrect model, which is called the *Bayes risk*, given by

$$p_{\mathrm{err}} \triangleq \sum_i \sum_{j \neq i} \int_{\mathcal{R}_j} p_{\mathbf{z}|\mathbf{H}}(z|H_i) p(H_i) dz \tag{16}$$

where $\mathcal{R}_j \triangleq \{z \in \mathcal{Z} : p_{\mathbf{H}|\mathbf{z}}(H_j|z) > p_{\mathbf{H}|\mathbf{z}}(H_k|z), \forall k \neq j\}$ is the region in which the hypothesis H_j is the most probable. Computing the exact p_{err} requires high computational demand that corresponds to sums of many multi-dimensional integrals. Computation can be relaxed by using the Bhattacharyya bound that provides an upper bound on the Bayes risk and is defined as

$$\bar{e}_{\mathrm{Bhat}}(H_i, H_j) \triangleq p_{ij} \int \sqrt{p_{\mathbf{z}|\mathbf{H}}(z|H_i) p_{\mathbf{z}|\mathbf{H}}(z|H_j)} dz \tag{17}$$

where $p_{ij} \triangleq \sqrt{p(H_i)p(H_j)}$ are constants related to the prior distributions of the hypotheses H_i and H_j.

Define a distance measure between two probability distributions $f = \mathcal{N}(\mu_1, \Sigma_1)$ and $g = \mathcal{N}(\mu_2, \Sigma_2)$ by

$$\begin{aligned} \rho_{\mathrm{Bhat}}(f, g) \triangleq \frac{1}{2}\Bigg(&\ln \frac{\det(\Sigma_1 + \Sigma_2)}{\det \Sigma_1 \det \Sigma_2} + n \ln \frac{1}{2} \\ &+ \frac{1}{2}(\mu_1 - \mu_2)^{\mathrm{T}} (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)\Bigg). \end{aligned} \tag{18}$$

Consider the set of linear hypothesized models $\mathcal{M}$ given in (5) that are associated with the set of hypotheses $\mathcal{H}$. Then, the Bhattacharyya bound (17) has the following relation with the distance measure given in (18):

$$\ln \bar{e}_{\mathrm{Bhat}}(H_i, H_j) \leq \ln p_{ij} - \rho_{\mathrm{Bhat}}(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j))).$$

V. OPTIMAL INPUT DESIGN FOR FDD: CONSTRAINED OPTIMIZATION AND CONVEX RELAXATION

There can be situations when two or more hypotheses are nearly equally probable and so are not distinguishable from the current observable data because their predicted (hypothesized) distributions are quantitatively very close. To resolve such difficult decision-making situations, we consider optimal input design problems for which the control input maximizing detectability of faults is constructed while retaining desirable system behaviors or minimizing degradation of system performance incurred by FDD. Most of the existing fault diagnosis methods are *passive* in the sense that those diagnostic procedures are based on the observed data for given inputs. The input design for fault diagnosis considered here is an active approach to facilitate statistical decision location of the true fault. For two different models corresponding to different faults, sensitivity of the observables' statistics to input changes can be substantially different from each other. The inputs are changed within an allowable range of operation so that the resultant statistics of observables predicted from different fault scenarios are notably different and the more probable fault scenario obtained from hypothesis testing is diagnosed as a most likely estimated fault. The distinguishability between two models of fault scenarios can be quantified by

$$\begin{aligned} \delta_{ij}(z) \triangleq \rho(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j)) = \delta_{ji}(z) \text{ or} \\ \delta_{ij}(z) \triangleq \rho(p_{\mathbf{H}|\mathbf{z}}(H_i|z), p_{\mathbf{H}|\mathbf{z}}(H_j|z)) = \delta_{ji}(z) \end{aligned} \tag{19}$$

where $\rho(\cdot, \cdot)$ refers to a certain (symmetric) measure of distance between two probability distributions presented in Section IV-A.

A. Constraints on Predicted Controlled Trajectories

The constraints $\mathcal{U}$ and $\mathcal{Y}$ are assumed to be convex and several classes of those constraints considered here are

$$\begin{aligned} \mathcal{U}_1(\gamma_1^u) &\triangleq \{u : \|u\|_1 \leq \gamma_1^u\}, \\ \mathcal{U}_\infty(u_{\min}, u_{\max}) &\triangleq \{u : u_{\min} \leq u \leq u_{\max}\}, \\ \mathcal{U}_2(\gamma_{\mathrm{p}}^u) &\triangleq \{u : \|u\|_2 \leq \gamma_{\mathrm{p}}^u\}, \end{aligned} \tag{20}$$

and

$$\mathcal{Y}_\infty(y_{\min}, y_{\max}) \triangleq \{y : y_{\min} \leq y \leq y_{\max}\}, \tag{21}$$

The constraint sets $\mathcal{U}$ and $\mathcal{Y}$ can also be intersections of the constraints in (20) and (21), respectively.

B. Constrained Optimization for Model Discrimination

Consider two models of faults M_i and M_j corresponding to hypotheses H_i and H_j. The goal is to find optimum data z that solve the constrained optimization

$$\max_{z \in \mathcal{U} \times \mathcal{Y}} \rho(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j)) \tag{22}$$

where $\mathcal{U}$ and $\mathcal{Y}$ refer to the input and output constraints, respectively, given in Section V-A.

Lemma 1. Suppose that the control input u is independent of the output y, i.e., u is an open-loop control. Then

$\rho(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j))$ is convex in z, or equivalently in u, for any distance measure $\rho \in \{\rho_{\mathrm{KL}}^{\min}, \rho_{\mathrm{KL}}^{\mathrm{ave}}, \rho_{\mathrm{Bhat}}\}$.

C. Multi-step Lookahead Maximization of Distinguishability between Two Competing Hypothesized Models

Consider the LTI system models (5). Then the concatenated output trajectory within the time interval of prediction horizon $[\kappa+1, \kappa+m_h]$ is

$$\mathbf{y}_{\kappa,m_h}^i = \mathcal{F}_u^i \mathbf{u}_{\kappa,m_h} + \mathcal{F}_x^i x_\kappa + \mathcal{F}_w^i \mathbf{w}_{\kappa,m_h}^i + \mathcal{F}_v^i \mathbf{v}_{\kappa,m_h}^i \tag{23}$$

where $\mathbf{y}_{\kappa,m_h}^i = y_{\kappa+1:\kappa+m_h}^i$, $\mathbf{u}_{\kappa,m_h} = u_{\kappa:\kappa+m_h-1}$, $\mathbf{w}_{\kappa,m_h}^i = w_{\kappa:\kappa+m_h-1}^i$, $\mathbf{v}_{\kappa,m_h}^i = v_{\kappa+1:\kappa+m_h}^i$.

The corresponding mean and covariance of the predicted controlled output trajectory are given by

$$\mu_{\kappa,m_h}^i \triangleq \mathbf{E}[\mathbf{y}_{\kappa,m_h}^i] \tag{24}$$

and

$$\Sigma_{\kappa,m_h}^i \triangleq \mathbf{E}\big[(\mathbf{y}_{\kappa,m_h}^i - \mu_{\kappa,m_h}^i)(\mathbf{y}_{\kappa,m_h}^i - \mu_{\kappa,m_h}^i)^{\mathrm{T}}\big]\,. \tag{25}$$

The covariance is independent of the control input sequence $\mathbf{u}_{\kappa,m_h}$ provided that it is open-loop control. For notational convenience, rewrite

$$\mu_{\kappa,m_h}^i = \mathcal{G}_u^i \mathbf{u}_{\kappa,m_h} + g_\kappa^i. \tag{26}$$

For optimality criteria for the control input maximizing the statistical distance between two hypothesized system models, consider

$$J_{ij}(\mathbf{u}_{\kappa,m_h};\rho) = (\tilde{\mu}_{\kappa,m_h}^{ij})^{\mathrm{T}} \mathcal{P}_\rho^{ij} (\tilde{\mu}_{\kappa,m_h}^{ij}) \tag{27}$$

where $\tilde{\mu}_{\kappa,m_h}^{ij} \triangleq \mu_{\kappa,m_h}^i - \mu_{\kappa,m_h}^j$ and each $\mathcal{P}_\rho^{ij}$ is defined by one of the positive-definite matrices:

$$\begin{cases} (\Sigma_{\kappa,m_h}^i)^{-1} + (\Sigma_{\kappa,m_h}^j)^{-1} & \text{for } \rho = \rho_{\mathrm{KL}}^{\mathrm{ave}}, \\ (\Sigma_{\kappa,m_h}^i + \Sigma_{\kappa,m_h}^j)^{-1} & \text{for } \rho = \rho_{\mathrm{Bhat}}. \end{cases} \tag{28}$$

Similarly, consider

$$\begin{aligned} J_{ij}(\mathbf{u}_{\kappa,m_h};\rho_{\mathrm{KL}}^{\max}) &= \max_{k\in\{i,j\}} (\tilde{\mu}_{\kappa,m_h}^{ij})^{\mathrm{T}} \Sigma_{\kappa,m_h}^k (\tilde{\mu}_{\kappa,m_h}^{ij}), \\ J_{ij}(\mathbf{u}_{\kappa,m_h};\rho_{\mathrm{KL}}^{\min}) &= \min_{k\in\{i,j\}} (\tilde{\mu}_{\kappa,m_h}^{ij})^{\mathrm{T}} \Sigma_{\kappa,m_h}^k (\tilde{\mu}_{\kappa,m_h}^{ij}), \end{aligned} \tag{29}$$

which can be rewritten as the quadratic form

$$J(\mathbf{u}_{\kappa,m_h};\rho) \triangleq \mathbf{u}_{\kappa,m_h}^{\mathrm{T}} \mathcal{Q}_\rho \mathbf{u}_{\kappa,m_h} + q_\rho^{\mathrm{T}} \mathbf{u}_{\kappa,m_h} + q_{\rho,0} \tag{30}$$

where the super- and subscripts ij are dropped due to simplify notation. With this general form of optimality measure, the resultant constrained optimization can be represented as

$$\begin{aligned} \min_{\mathbf{u}_{\kappa,m_h}} \quad & -J(\mathbf{u}_{\kappa,m_h};\rho) \\ \text{s.t.} \quad & \mathbf{u}_{\kappa,m_h} \in \mathcal{U}, \\ & \mathcal{G}_u^i \mathbf{u}_{\kappa,m_h} + g_\kappa^i \in \mathcal{Y}, \end{aligned} \tag{31}$$

where the output constraints are imposed on the expected output trajectory within the prediction horizon interval $[\kappa+1, \kappa+m_h]$. The symmetric matrix $\mathcal{Q}_\rho$ in $J(\mathbf{u}_{\kappa,m_h};\rho)$ is positive definite for all ρ under consideration, which implies that the optimization problem (31) is nonconvex since the objective function is concave in $\mathbf{u}_{\kappa,m_h}$.

D. Semidefinite Relaxation

The optimization problem (31) can be rewritten as

$$\begin{aligned} \min_{\mathbf{u}_{\kappa,m_h}} \quad & -\mathrm{Tr}(\mathcal{Q}_\rho \mathbf{U}_{\kappa,m_h}) - q_\rho^{\mathrm{T}} \mathbf{u}_{\kappa,m_h} - q_{\rho,0} \\ \text{s.t.} \quad & \mathbf{U}_{\kappa,m_h} = \mathbf{u}_{\kappa,m_h} \mathbf{u}_{\kappa,m_h}^{\mathrm{T}},\ \mathbf{u}_{\kappa,m_h} \in \mathcal{U}, \\ & \mathcal{G}_u^k \mathbf{u}_{\kappa,m_h} + g_\kappa^k \in \mathcal{Y},\ k \in \{i,j\}, \end{aligned} \tag{32}$$

where a dummy matrix variable $\mathbf{U}_{\kappa,m_h}$ is introduced and the first equality corresponds to the only nonconvex relation. More general and explicit form of the optimization can be written as a (nonconvex) QCQP

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}^{\mathrm{T}} \mathcal{Q} \mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^{\mathrm{T}} \mathcal{A}_\ell \mathbf{x} \geq 0,\ \ell = 1,\ldots,m_q, \\ & \mathcal{B}\mathbf{x} \geq 0, \\ & \mathbf{x} = [1\ \mathbf{u}_{\kappa,m_h}^{\mathrm{T}}]^{\mathrm{T}}, \end{aligned} \tag{33}$$

where $\mathcal{Q} \triangleq \begin{bmatrix} -q_{\rho,0} & -1/2q_\rho^{\mathrm{T}} \\ -1/2q_\rho & -\mathcal{Q}_\rho \end{bmatrix}$, and the matrices $\mathcal{A}_\ell$ and $\mathcal{B}$ can be explicitly obtained from the constraint sets $\mathcal{U}$ and $\mathcal{Y}$ and the system matrices associated with the hypothesized models indexed by $k \in \{i,j\}$. This (nonconvex) QCQP can be rewritten as

$$\begin{aligned} \min_{\mathbf{X}} \quad & \mathrm{Tr}(\mathcal{Q}\mathbf{X}) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathcal{A}_\ell \mathbf{X}) \geq 0,\ \ell = 1,\ldots,m_q, \\ & \mathcal{B}\mathbf{X}e_1 \geq 0, \\ & \mathcal{B}\mathbf{X}\mathcal{B}^{\mathrm{T}} \geq 0, \\ & e_1^{\mathrm{T}} \mathbf{X} e_1 = 1, \\ & \mathbf{X} \succeq 0, \\ & \mathrm{rank}(\mathbf{X}) = 1, \end{aligned} \tag{34}$$

where e_1 denotes the first standard basis vector in $\mathbb{R}^{m_h+1}$.

By removing the (nonconvex) rank constraint, $\mathrm{rank}(\mathbf{X}) = 1$, the corresponding primal SDP relaxation is

$$\begin{aligned} \min_{\mathbf{X}} \quad & \mathrm{Tr}(\mathcal{Q}\mathbf{X}) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathcal{A}_\ell \mathbf{X}) \geq 0,\ \ell = 1,\ldots,m_q, \\ & \mathcal{B}\mathbf{X}e_1 \geq 0, \\ & \mathcal{B}\mathbf{X}\mathcal{B}^{\mathrm{T}} \succeq 0, \\ & e_1^{\mathrm{T}} \mathbf{X} e_1 = 1, \\ & \mathbf{X} \succeq 0, \end{aligned} \tag{35}$$

where e_1 denotes the first standard basis vector in $\mathbb{R}^{m_h+1}$.

Using different measures defined in (29), the objective function in the SDP relaxation (35) can be replaced by $\min_{\mathbf{X}} \max_{k\in\{i,j\}} \mathrm{Tr}(\mathcal{Q}^k\mathbf{X})$ or $\min_{\mathbf{X}} \min_{k\in\{i,j\}} \mathrm{Tr}(\mathcal{Q}^k\mathbf{X})$ where the symmetric matrices $\mathcal{Q}^k$ can be computed from (29) and the associated system matrices for the hypothesized models indexed by $k \in \{i,j\}$. Notice that $\max_{k\in\{i,j\}} \mathrm{Tr}(\mathcal{Q}^k\mathbf{X})$ is convex whereas $\min_{k\in\{i,j\}} \mathrm{Tr}(\mathcal{Q}^k\mathbf{X})$ is concave, which indicates a preference for $\rho_{\mathrm{KL}}^{\min}$ instead of $\rho_{\mathrm{KL}}^{\max}$.

Lemma 2 (*[19]*). Consider a hypercube constraint $\mathcal{U} = \mathcal{U}_\infty(\cdot,\cdot)$ and $\mathcal{Y} = \mathbb{R}^{N_h}$. The performance bounds achieved by the SDP relaxation is

$$J_{\mathrm{sdp}}^\star \leq J_{\mathrm{qcqp}}^\star \leq \frac{\pi}{2} J_{\mathrm{sdp}}^\star.$$

E. Optimality Criteria for Model Discrimination with Multiple Hypotheses

Suppose that there are more than two candidate fault scenarios including the nominal operation, $\mathcal{H}_N \triangleq \{H_1, \cdots, H_N\}$. For an optimality criterion that quantifies information included in the predicted input-output data $z \in \mathcal{U} \times \mathcal{Y}$, we propose to consider one of the following:

i. Maximizing the minimum statistical distance among the hypothesized models

$$\max_{z \in \mathcal{U} \times \mathcal{Y}} \min_{\{j>i, i=1,\dots,N\}} \rho_{ij}(z); \tag{36}$$

ii. Maximizing the average statistical distance of the hypothesized models

$$\max_{z \in \mathcal{U} \times \mathcal{Y}} \sum_{i=1}^{N-1} \sum_{j>i}^{N} \rho_{ij}(z); \tag{37}$$

iii. Maximizing the weighted average statistical distance of the hypothesized models

$$\max_{z \in \mathcal{U} \times \mathcal{Y}} \sum_{i=1}^{N-1} \sum_{j>i}^{N} \gamma_{ij} \rho_{ij}(z),\ \gamma_{ij} \in [0,1],\ \sum_{i=1}^{N-1} \sum_{j>i}^{N} \gamma_{ij} = 1, \tag{38}$$

where $\rho_{ij}(z) \triangleq \rho(p_{\mathbf{z}|\mathbf{H}}(z|H_i), p_{\mathbf{z}|\mathbf{H}}(z|H_j))$ with $\rho \in \{\rho_{\text{KL}}^{\text{min}}, \rho_{\text{KL}}^{\text{ave}}, \rho_{\text{Bhat}}\}$.

Remark 1. Note that if $\rho_{ij}(\cdot)$ are concave for all indices then the aforementioned optimizations (36), (37), and (38) are all convex, provided that the constraint $\mathcal{U} \times \mathcal{Y}$ is convex.

F. Randomized Algorithms: Suboptimal Solutions

A randomized algorithm for computing a rank 1 solution from the optimal solutions $\mathbf{X}_\ell^\star$, $\ell = 1, \dots, N$, of the SDP relaxation (35) is presented in [19], [20]. Using a Cholesky factorization $P\mathbf{X}_\ell^\star P^{\mathrm{T}} = S_\ell^{\mathrm{T}} S_\ell$,‡

$$\tilde{\mathbf{x}}_\ell := D\text{sgn}(\mathrm{S}_\ell^{\mathrm{T}} \xi),$$

where ξ is a Gaussian random vector whose distribution is $\mathcal{N}(0, \mathrm{I})$ and $D > 0$ is the diagonal scaling matrix such that $\tilde{\mathbf{x}}$ satisfies the constraints in (33). An optimal scaling matrix may be defined by the convex optimization

$$\begin{aligned} &\max \ \|\text{diag}(D)\|_p \\ &\text{s.t. } \tilde{\mathbf{x}}_\ell = D\text{sgn}(\mathrm{S}_\ell^{\mathrm{T}} \xi),\ \tilde{\mathbf{x}}_\ell \in \mathcal{C} \end{aligned} \tag{39}$$

where $p \in [1, \infty]$ refers in the vector p-norm, $\mathcal{C}$ is the intersection of the quadratic and linear constraints given in (33), and ξ is a realization from the distribution $\mathcal{N}(0, \mathrm{I})$. Another method for computing a rank 1 solution from the optimal solutions $\mathbf{X}_\ell^\star$ of the SDP relaxation (35) is a biased randomized algorithm:

$$\tilde{\mathbf{x}}_\ell := \bar{\mathbf{x}}_\ell + D\text{sgn}(\mathrm{S}_\ell^{\mathrm{T}} \xi)$$

where $\bar{\mathbf{x}}_\ell$ is the singular vector corresponding to the largest singular value of $P\mathbf{X}_\ell^\star P^{\mathrm{T}}$. Another method is to compute

$$\tilde{\mathbf{x}}_\ell := \bar{\mathbf{x}}_\ell' + D\text{sgn}(\mathrm{S}_\ell^{\mathrm{T}} \xi)$$

where $\bar{\mathbf{x}}_\ell' = (\mathbf{X}_\ell^\star)_{2:N_h,1}$ is the vector obtained from the first column vector of $\mathbf{X}_\ell^\star$ from which the first element is excluded.

By generating N_{s} samples $\{\xi_n\}_{n=1}^{N_{\text{s}}}$ of ξ from the distribution $\mathcal{N}(0, \mathrm{I})$, compute an approximate suboptimal solution

$$\hat{\mathbf{x}} := \arg \min_{n=1,\cdots,N_{\text{s}}} \begin{bmatrix} 1 \\ \tilde{\mathbf{x}}_n \end{bmatrix}^{\mathrm{T}} \mathcal{Q} \begin{bmatrix} 1 \\ \tilde{\mathbf{x}}_n \end{bmatrix} \tag{40}$$

where $\tilde{\mathbf{x}}_n$ is a feasible solution for the constraints in (33) associated with the nth sample ξ_n. The notation $\hat{\mathbf{x}}(N_{\text{s}})$ can be used to denote its dependence on the number of samples.

VI. DISCUSSION

A. Simulation Results

To illustrate and compare the input design methods, consider fault scenarios for an aircraft system. The numerical data are adopted from [6]. Consider a discretized dynamical system model for the longitudinal aircraft dynamics, in which the sampling time is 0.5 sec. The state variables are $x = [V_y, V_x, \omega, \theta, L]^{\mathrm{T}}$ where V_y, V_x, $\omega = \dot{\theta}$, θ, and L refer to the vertical velocity, horizontal velocity, pitch rate, pitch angle, and altitude, respectively. The measurable outputs are $y = [V_y, \omega]^{\mathrm{T}}$. For nominal operation without any faults corresponding to the hypothesis H_0, the resulting system matrices are

$$A^0 = \begin{bmatrix} 0.9985 & 0.1950 & 0 & -0.161 & 0 \\ -0.0325 & 0.8405 & 3.87 & 0 & 0 \\ 0.01 & -0.0505 & 0.7855 & 0 & 0 \\ 0 & 0 & 0.5 & 1.0 & 0 \\ 0.5 & 0 & 0 & 0 & 1.0 \end{bmatrix},$$

$$B^0 = \begin{bmatrix} 0.005 \\ -0.09 \\ -0.58 \\ 0 \\ 0 \end{bmatrix},\ E^0 = \mathrm{I}_5,$$

$$C^0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix},\ D^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ F^0 = \mathrm{I}_2.$$

Consider three different types of fault scenarios of the hypotheses and the associated system matrices given by the followings:

- ○ H_1: Failure of the vertical velocity sensor
  - $M_1 = (A^1, B^1, C^1, D^1, E^1, F^1)$ where $A^1 = A^0$, $B^1 = B^0$, $D^1 = D^0$, $E^1 = E^0$, $F^1 = F^0$, and $C^1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$;
- ○ H_2: Failure of the pitch rate sensor
  - $M_2 = (A^2, B^2, C^2, D^2, E^2, F^2)$ where $A^2 = A^0$, $B^2 = B^0$, $D^2 = D^0$, $E^2 = E^0$, $F^2 = F^0$, and $C^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$;
- ○ H_3: Failure of the elevator actuator
  - $M_3 = (A^3, B^3, C^3, D^3, E^3, F^3)$ where $A^3 = A^0$, $C^3 = C^0$, $D^3 = D^0$, $E^3 = E^0$, $F^3 = F^0$, and $B^3 = 0_{5\times 1}$.

The inputs are assumed to be bounded in amplitude, $\mathcal{U}_\infty(-1/2, 1/2)$.

Example 1. Consider $\mathcal{H} = \{H_0, H_1, H_2, H_3\}$. Fig. 1 shows the input trajectories computed from solving the associated convex relaxations of optimal input design problems with different statistical distance measures. Since there are more than two models of hypotheses to compare, we solve a minimax problem for which an optimal solution minimizes the

---

‡ $P\mathbf{X}_\ell^\star P^{\mathrm{T}}$ is the matrix obtained by removing the first row and column of $\mathbf{X}_\ell^\star$.

maximum among the statistical distance measures of each pair of hypotheses. The number of pairs is $N(N+1)/2$ where N is the number of system models (or modes) for hypotheses, which implies that the associated convex relaxation for a multiple hypotheses test can be still solved in polynomial-time. Figs. 2 and 3 present the resulting trajectories of V_y and ω, respectively, in which sharp distinctions between the four models can be observed for all three input design methods. The output V_y shows the distinct behaviors for the models M_0 and M_2 whereas the models M_1 and M_3 have the same trajectory. The output ω shows the different behaviors for the models M_1 and M_3 whereas the models M_0 and M_2 have the same trajectory.

VII. CONCLUSION

This paper presents semidefinite programming relaxations of optimal *active* input design problems for fault detection and diagnosis. For an FDD procedure based on statistical decision theory (in particular, Bayesian inference), the resulting optimization is to maximize statistical discrimination between hypothesized fault scenarios. Each fault scenario model characterizes a random process of the measurable outputs; to quantify the quality of the measurable data for FDD, this paper considers three different measures for the statistical distance between two random processes. With such statistical distance measures, the original optimization is nonconvex even without any constraints, which would be computationally expensive, especially for multiple hypothesis tests. Convex relaxation methods are proposed that compute approximate solutions for which the potential degree of suboptimality is known in some special cases. Simulation results are included to demonstrate the active input design methods for FDD.

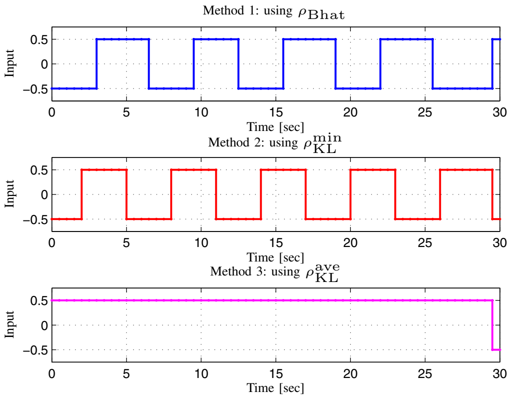

Fig. 1: *Input sequences obtained from the design methods for multiple hypotheses* $\mathcal{H} = \{H_0, H_1, H_2, H_3\}$.

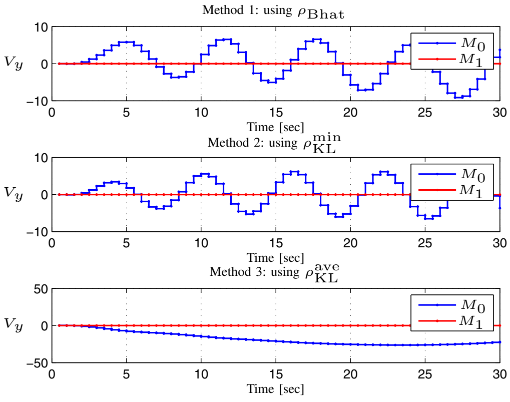

Fig. 2: *The expected trajectory of* V_y *generated by the input design methods for multiple hypotheses* $\mathcal{H} = \{H_0, H_1, H_2, H_3\}$.

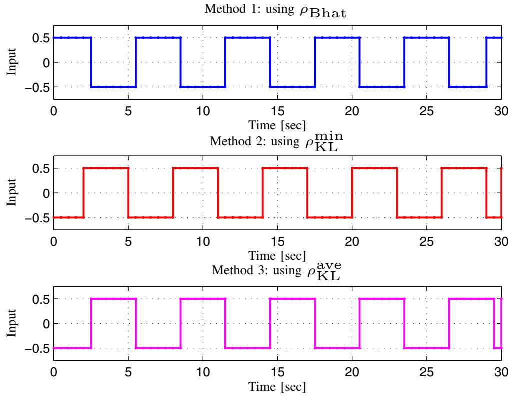

Fig. 3: *The expected trajectory of* ω *generated by the presented design methods for multiple hypotheses* $\mathcal{H} = \{H_0, H_1, H_2, H_3\}$.

REFERENCES

[1] X. J. Zhang, "Auxiliary signal design in fault detection and diagnosis," in *Lecture Notes in Control and Information Sciences, Vol. 134*. Berlin: Springer-Verlag, 1989.
[2] F. Kerestecioglu and M. B. Zarrop, "Input design for detection of abrupt changes in dynamical systems," *Int. J. of Control*, vol. 59, pp. 1063–1084, 1994.
[3] S. L. Campbell and R. Nikoukhah, *Auxiliary Signal Design for Failure Fetection*. Princeton, NJ: Princeton University Press, 2004.
[4] H. H. Niemann, "A setup for active fault diagnosis," *IEEE Trans. on Auto. Contr.*, vol. 51, pp. 1572–1578, 2006.
[5] M. Simandl and I. Puncochar, "Active fault detection and control: unified formulation and optimal design," *Automatica*, vol. 45, pp. 2052–2059, 2009.
[6] L. Blackmore, S. Rajamanoharan, and B. Williams, "Active estimation for jump Markov linear systems," *IEEE Trans. on Auto. Contr.*, vol. 53, pp. 2223–2236, 2008.
[7] H. H. Niemann and J. Stoustrup, "Robust fault detection in open loop vs. closed loop," in *IEEE CDC*, 1997, pp. 4496–4497.
[8] R. Nikoukhah, S. L. Campbell, K. G. Horton, and F. Delebecque, "Auxiliary signal design for robust multimodel identification," *IEEE Trans. on Auto. Contr.*, vol. 47, pp. 158–164, 2002.
[9] M. Simandl, I. Puncochar, and J. Kralovec, "Rolling horizon for active fault detection," in *IEEE CDC*, 2005, pp. 3789–3794.
[10] A. E. Ashari, R. Nikoukhah, and S. Campbell, "Effects of feedback on active fault detection," *Automatica*, vol. 48, pp. 866–872, 2012.
[11] ——, "Active robust fault detection in closed-loop systems: Qquadratic optimization approach," *IEEE Trans. on Auto. Contr.*, vol. 54, pp. –, 2012, in press.
[12] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. UK: Cambridge University Press, 2003.
[13] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Motivations to Computational Implementation*. NY: Springer-Verlag, 2001.
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: John Wiely & Sons, Inc., 2006.
[15] Q. Zheng, "A volterra series approach to nonlinear process control and control-relevant identification," Ph.D. dissertation, Univ. of Maryland, College Park, 1995.
[16] A. P. Featherstone and R. D. Braatz, "Input design for large scale sheet and film processes," *Ind. Eng. Chem. Res.*, vol. 37, pp. 449–454, 1998.
[17] B. Cooley, J. Lee, and S. Boyd, "Control-relevant experiment design: A plant-friendly, lmi-based approach," in *Proc. of ACC*, 1998, pp. 1240–1244.
[18] D. Rivera, H. Lee, M. Braun, and H. Mittelmann, "Plant friendly system identification: A challenge for the process industries," in *IFAC SYSID*, 2003, pp. 917–922.
[19] Y. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optimization Methods and Software*, vol. 9, pp. 141–160, 1998.
[20] Y. Nesterov, H. Wolkowicz, and Y. Ye, "Nonconvex quadratic optimization," in *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, H. Wolkowicz, R. Saigal, and L. Vandenberghe, Eds. Boston, MA: Kluwer Academic Publishers, 2000.