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

Optimum Input Design for Fault Detection and Diagnosis: Model-based Prediction and Statistical Distance Measures

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

***Abstract*— This paper proposes optimization-based *active* fault detection and diagnosis (FDD) methods. An optimal input sequence is computed for maximizing discrimination between system models of fault scenarios in a statistical sense. Two different measures quantifying the degree of distinguishability between two stochastic LTI system models are considered, and their geometric properties are investigated. Their connection to the generalized likelihood ratio tests are also presented. Constrained open- and closed-loop feedback input design methods using model-based prediction are presented. Constraints on the predicted controlled output trajectory are imposed for ensuring operational safety as well as the input constraints that correspond to hardware limitations. Receding horizon method is used to implement the computed inputs.**

I. Introduction

As systems become more complicated, faults become inevitable and the purpose of a fault detection and diagnosis (FDD) algorithm is to determine whether and where a fault occurs so that these faults can be corrected in a timely manner. Integrating FDD and feedback control algorithms allows to increase overall closed-loop performance, for example, by enabling the closed-loop system to function during the occurrence of faults as long as the overall system is still reasonably controllable. The closed-loop performance of model-based controllers degrade in the presence of plant-model mismatches corresponding to faults and this performance can be significantly improved if the control system is fed information on the fault condition from the FDD algorithm. In particular, the control system can be reconfigured in response to the fault condition. Such change detection is related to hybrid automata and mode estimation [1], [2].

Note that FDD could be particularly difficult in presence of feedback control because this latter can mask the ability of the FDD algorithm to detect and diagnose faults. For this reason, this paper considers *active* FDD approaches that excite or perturb the manipulated variables to improve the detection and diagnosis of faults [3]–[8]. Many stochastic [9] and deterministic [10]–[12] methods for the design of such auxiliary inputs have been proposed. A main difference between stochastic and deterministic approaches is in the representation of uncertain system models for which stochastic approaches use stochastic differential equations (SDEs) and deterministic approaches use set-valued operators defined over compact supports. Some researchers have investigated the effects of feedback on performance of FDD and quadratic cost optimality criteria [13], [14], and others have considered finite- or infinite-horizon control methods [7], [9], [12].

In this paper, we consider linear stochastic dynamical systems with additive Gaussian disturbance and noise, so that models of fault scenarios define Gaussian random processes. To quantify statistical distance between two Gaussian processes, the well-known KL-divergence and an associated approximation using geometric properties of Gaussian distributions are used. In particular, this paper (i) studies the properties of the two aforementioned measures of statistical distance between probability distributions associated with two different fault scenario models and (ii) proposes methods for the design of inputs for *robust active* FDD in terms of solutions of mathematical programs for which a sequential method of semidefinite programs (SDPs) is applied. It is observed that the approximate distance measure using geometric properties of Gaussian distributions can be compatible with a closed-loop static state feedback controller that might outperform open-loop input design methods, whereas the direct use of the KL-divergence might not be trivial. As theoretical contributions for the problem formulation of optimum input design for FDD, we investigate the relations of the proposed measures of statistical distance to the generalized likelihood ratio tests and justify the use of such measures of statistical distance for quantifying the degree of discrimination between two hypothesized models.

This paper develops a unified and systematic framework for optimal probing input design for maximizing performance of statistical FDD algorithms that incorporates information theoretic measures of differences between stochastic dynamical systems. A sequential method of SDP is proposed for solving non-convex constrained optimization problems from which local optimal solutions can be obtained. To further improve performance of FDD, a design method of closed-loop state feedback control is proposed, in which the control gain is separately designed from the affine input. An underlying assumption is that the FDD procedure is based on a statistical decision that solves Bayesian inference problems for which the measurements 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 confidence in 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 is to facilitate the associated statistical decision and maximize its robustness.

Notation: The following notation is used: $\mathbf{E}$ or $\bar{(\cdot)}$ is the expectation or mean; $\mathbf{Var}$ or $\Sigma_{(\cdot)}$ is the variance or covariance; $\mathcal{N}(a,b)$ is the Gaussian distribution with the mean a and the covariance b; the symbol $\sim$ means "distributed as"; $\mathbb{S}^n \subset \mathbb{R}^{n\times n}$ is the set of real symmetric matrices; and its subsets $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ denote the set of positive semidefinite and definite matrices, respectively. An equivalent nomenclature is to use $X \succeq 0$ ($X \succ 0$) to represent $X \in \mathbb{S}^n_+$ ($X \in \mathbb{S}^n_{++}$).

K.-K. K. Kim is with the University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA, kkim78@illinois.edu.

D. M. Raimondo is with the Universita' degli Studi di Pavia, Pavia, Italy, davide.raimondo@unipv.it.

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

978-3-952-41734-8/©2013 EUCA

II. Preliminaries: Bayesian Hypothesis Tests

This section provides a brief review on Bayesian hypothesis testing (see [15], [16] for more details). 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 that explains 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 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

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

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 $C : \mathcal{H} \times \mathcal{H} \to \mathbb{R}_+$, $C(H, h(\cdot))$ is the cost of deciding that the hypothesis is h 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. Due to linearity and whiteness of the Gaussian random processes w and v, the corresponding random processes x_t and y_t solving SDE (4) have jointly Gaussian distributions.

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

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

where the superscript i refers to the occurrence of the i^{th} hypothesized fault. The fault detection and diagnosis procedure 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. We also note that the hypothesized models can have different dimensions of the state variables, whereas the number of outputs are the same.

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. Roughly speaking, the further statistical distance between two hypotheses, the lower probability of false alarm.

1) Relative Entropy as a Distance Measure: A common measure of the distance between two probability distributions is relative entropy, which is also called the Kullback-Leibler distance (or divergence).

Definition 1 (*See also [17] for details*). For two probability density functions f and g, the *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, the 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, the relative entropy is defined by (6) and its value is 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 a symmetric distance measure, define

$$\rho_{\mathrm{KL}}(f, g) \triangleq \min\{d_{\mathrm{KL}}(f||g), d_{\mathrm{KL}}(g||f)\} \tag{8}$$

such that $\rho_{\mathrm{KL}}(f, g) = \rho_{\mathrm{KL}}(g, f)$. Consider two different Gaussian distribution functions $f : \mathbb{R}^n \to \mathbb{R}_+$ and $g : \mathbb{R}^n \to \mathbb{R}_+$. In particular, for two independent Gaussian distributions $f = \mathcal{N}(\mu_1, \Sigma_1)$ and $g = \mathcal{N}(\mu_2, \Sigma_2)$, the KL distances are given by

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

and

$$\begin{aligned} d_{\mathrm{KL}}(g||f) = \tfrac{1}{2} \big(& \ln\det\Sigma_1 - \ln\det\Sigma_2 + \mathrm{Tr}\,\Sigma_1^{-1}\Sigma_2 \\ &+ (\mu_1 - \mu_2)^{\mathrm{T}} \Sigma_1^{-1} (\mu_1 - \mu_2) - n \big), \end{aligned} \tag{10}$$

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

*Definitions for the relative entropy of two probability distributions of different types of supports, i.e., continuous and discrete support sets, are not studied in classical information theory. However, the relative entropy is defined with a measure function $f(\cdot)$ for both 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 set.

2) Geometric Distance Measure Between Two Gaussian Distributions: Consider two ellipsoids $\mathcal{E}_1(\mu_1, \Sigma_1, \gamma) \triangleq \{x \in \mathbb{R}^n : (x-\mu_1)^{\mathrm{T}}\Sigma_1^{-1}(x-\mu_1) \leq \gamma\}$ and $\mathcal{E}_2(\mu_2, \Sigma_2, \gamma) \triangleq \{x \in \mathbb{R}^n : (x-\mu_2)^{\mathrm{T}}\Sigma_2^{-1}(x-\mu_2) \leq \gamma\}$, where the positive constant γ corresponds to the scaling factor of the volume of the corresponding ellipsoid.

To quantify the statistical distance between two Gaussian distributions $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$, compute the largest value of $\gamma > 0$ such that $\mathcal{E}_1(\mu_1, \Sigma_1, \gamma) \cap \mathcal{E}_2(\mu_2, \Sigma_2, \gamma) = \emptyset$, which can be formulated as the two-stage SDP

$$\begin{aligned}
&\max_{\gamma}\ \gamma \\
&\text{s.t.}\ \ 0 < \min_{t,x,y}\ \ t \\
&\qquad\quad \text{s.t.}\ \ \begin{bmatrix} \gamma & (x-\mu_1)^{\mathrm{T}} \\ (x-\mu_1) & \Sigma_1 \end{bmatrix} \succeq 0, \\
&\qquad\qquad\ \ \begin{bmatrix} \gamma & (y-\mu_2)^{\mathrm{T}} \\ (y-\mu_2) & \Sigma_2 \end{bmatrix} \succeq 0, \\
&\qquad\qquad\ \ \begin{bmatrix} t & (x-y)^{\mathrm{T}} \\ (x-y) & \mathrm{I} \end{bmatrix} \succeq 0.
\end{aligned} \tag{11}$$

For an alternative way to compute the largest value of $\gamma > 0$ such that $\mathcal{E}_1(\mu_1, \Sigma_1, \gamma) \cap \mathcal{E}_2(\mu_2, \Sigma_2, \gamma) = \emptyset$, consider the SDP

$$\begin{aligned}
&\min_{\gamma,x}\ \gamma \\
&\text{s.t.}\ \begin{bmatrix} \gamma & (x-\mu_1)^{\mathrm{T}} \\ (x-\mu_1) & \Sigma_1 \end{bmatrix} \succeq 0, \\
&\qquad \begin{bmatrix} \gamma & (x-\mu_2)^{\mathrm{T}} \\ (x-\mu_2) & \Sigma_2 \end{bmatrix} \succeq 0.
\end{aligned} \tag{12}$$

Then for any $\gamma < \gamma^*$ where γ^* is the optimal solution for (12), $\mathcal{E}_1(\mu_1, \Sigma_1, \gamma) \cap \mathcal{E}_2(\mu_2, \Sigma_2, \gamma) = \emptyset$. Furthermore, for $\Sigma_1, \Sigma_2 \in \mathbb{S}^n_{++}$, $\mathcal{E}_1(\mu_1, \Sigma_1, \gamma) \cap \mathcal{E}_2(\mu_2, \Sigma_2, \gamma)$ is a unique singleton. Define the distance measure for two Gaussian distributions f and g

$$\rho_{\mathrm{geo}}(f, g) \triangleq \gamma^* \tag{13}$$

where $f = \mathcal{N}(\mu_1, \Sigma_1)$, $g = \mathcal{N}(\mu_2, \Sigma_2)$, and γ^* is the optimal value of the SDP (12).

B. Gaussian m-array Hypothesis Testing for FDD

To formulate an optimization problem for Bayesian hypothesis testing, there are three elementary components that are required to be predetermined for multiple hypothesis tests (or m-array hypothesis tests) [15], [16]:

- m-array hypotheses with associated priors: Define m hypotheses. Without loss of generality, include the null hypothesis H_0 that corresponds to the normal operation, i.e., non-faulty model, for FDD. The total number of hypotheses to be tested are $m+1$. The corresponding a priori probability distributions are given by $P_i \triangleq \Pr[H = H_i]$.[†]
- Penalties for wrong decisions: Assign the penalty $C_{ij} \geq 0$ that corresponds to the cost to pay when the decision is $\hat{H} = H_i$, but the truth is $H = H_j$.
- Likelihood functions: Specify the closed-form of the propagation of hypothesis to the observables $p_{\mathbf{z}|\mathbf{H}}(z|H_i)$[‡] for each hypothesis H_i.

For FDD using a set of multiple models $\{M_0, \cdots, M_m\}$, each hypothesis is assigned to each associated model, i.e., $H = H_i \Leftrightarrow M = M_i$. For abuse of notation, $M = M_i$ is also used to refer to the associated hypothesis $H = H_i$.

The resulting optimization has the cost

$$\begin{aligned}
J(H_i, z) &= \sum_{j=0}^{m} C_{ij} \Pr[H = H_j | \mathbf{z} = z] \\
&= c \sum_{j=0}^{m} C_{ij} p_{\mathbf{z}|\mathbf{H}}(z|H_j) P_j
\end{aligned} \tag{14}$$

where $c \triangleq \sum_{j=0}^{m} p_{\mathbf{z}|\mathbf{H}}(z|H_j)P_j$ is the marginal probability that is independent of H_i and depends only on the (observable) data z. For a fixed realization z of the random variable $\mathbf{z}$, an optimal decision for hypothesis selection is

$$\hat{H}(z) = \arg \min_{H_i : i=0,\cdots,m} J(H_i, z), \tag{15}$$

which is a finite-state optimization for which a set of hypotheses are assessed and compared with each other. The required number of comparisons or tests increases as $O(m(m+1)/2)$ where m is the number of hypotheses to be assessed.

In dynamical systems, such Bayesian hypothesis testing is more likely comparing the predictions of the observables for multiple competing fault models (e.g., models given in (5)). For simplicity, assume that $z = y$, i.e., the measurement outputs are the only observables for hypothesis testing. The posterior distribution of the predicted output at time t with the system model M_k is

$$\begin{aligned}
\Pr[M = M_k, \eta_{0:t-1} | y_t] &= \Pr[y_t | \eta_{0:t-1}, M = M_k] \Pr[M_k] \\
&= \Pr[y_t | \eta_{t-1}, M = M_k] \Pr[M_k]
\end{aligned}$$

where $\eta \triangleq (x, u)$ and the last equality is due to the Markovian property induced by whiteness of process noise w_t and measurement noise v_t. Suppose that each hypothesized fault scenario has the system dynamics given by (5). Then the closed-forms of the likelihood functions $\Pr[y_t | \eta_{t-1}, M = M_k]$ can be computed, provided the previous system information compressed into η_{t-1} (or more precisely, its distribution $p(\eta_{t-1})$) can be accessed to every hypothesized model of a fault. In addition, we might assume reasonable accuracy of a state estimator.

To improve reliability of a Bayesian hypothesis testing, we can use a finite-time monitoring in which a finite sequence of the measurements is used to compute the posteriori distribution or more precisely to compute the likelihood function. Consider the monitoring window of length ℓ_m for which the sequence of the measurements $\{y_t, y_{t-1}, \cdots, y_{t-\ell_m+1}\}$ is monitored. The posterior distribution of the predicted measurements in this monitoring window with the system model M_k is

$$\begin{aligned}
&\Pr[M = M_k, x_{0:t-\ell_m}, u_{0:t-1} | y_{t-\ell_m+1:t}] \\
&\quad = \Pr[y_{t-\ell_m+1:t} | x_{0:t-\ell_m}, u_{0:t-1} M = M_k] \Pr[M_k] \\
&\quad = \Pr[y_{t-\ell_m+1:t} | x_{t-\ell_m}, u_{t-\ell_m:t-1}, M = M_k] \Pr[M_k].
\end{aligned}$$

[†]The choice of prior probability distributions might be essentially subjective and the *best* P_i is, in general, hard to determine in an objective way.

[‡]The observable vector z is assumed to consist of the variables that can be used for hypothesis testing, e.g., the inputs, measurements, and controlled outputs.

Assume that $\Pr[M_k] = 1/(m+1)$ for all $k = 0, \ldots, m$. Formally, the likelihood function for the k^{th} model M_k is defined as

$$\mathcal{L}_k(y) = \Pr[y_{t-\ell_m+1:t} = y | x_{t-\ell_m}, u_{t-\ell_m:t-1}, M = M_k] \tag{16}$$

where $y \in \mathbb{R}^{n_y \times \ell_m}$ refers to the measurements during the time interval corresponding to the monitoring window. Therefore,

$$k^\star(y) := \arg \max_{k=0,\cdots,m} \mathcal{L}_k(y)$$

where the argument y is explicitly represented to emphasize its dependence on the measurements that are indeed realizations of a random process in a finite interval.

Remark 1. When the monitoring window is moved forward, the previous (normalized) likelihood functions may be used as the prior distributions. In this case, there can be a longer delay in fault detection and diagnosis, but the probability of false alarm can decrease.

Remark 2. Consider the assumption that the initial condition at the starting point of the monitoring window and the applied control inputs are the same for all the system modes. Then the likelihood function can be rewritten as $\mathcal{L}_k(y; x_{t-\ell_m}, u_{t-\ell_m:t-1})$ and the corresponding optimal system mode rewritten as $k^\star(y; x_{t-\ell_m}, u_{t-\ell_m:t-1})$.

V. Optimal Input Design For FDD

It can occur that two or more hypotheses are almost equally probable 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, 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 *active* FDD. Many of the existing fault diagnosis methods are *passive* in the sense that those diagnostic procedures are based on the observed data for given inputs. Input design for fault diagnosis presented in this paper is an active approach to determine the true fault. For two different models corresponding to two different fault scenarios, the sensitivity of the observables' statistics to input changes can be substantially different from each other. Consider varying the inputs within an allowable range of operation so that the resultant statistics of observables predicted by different fault scenarios are notably different and the more probable fault scenario in a likelihood ratio hypothesis test is diagnosed as an estimated fault. A quantified measure of distinguishability between two models of fault scenarios is

$$\delta_{ij}(z) \triangleq \rho(p_{\mathbf{H}|\mathbf{z}}(H_i|z), p_{\mathbf{H}|\mathbf{z}}(H_j|z)) = \delta_{ji}(z)$$

where $\rho(\cdot,\cdot)$ denotes a certain (symmetric) measure of distance between two probability distributions. We consider the previously defined two measures of statistical distance ρ_{KL} and ρ_{geo} to quantify distinguishability of faults. Suppose that the input constraint $u \in \mathcal{U}$ is defined over a convex compact set $\mathcal{U}$ such as a polytope or ellipsoid.

A. Using ρ_{KL}

Consider the measure of statistical distance ρ_{KL} between two Gaussian distributions. Maximizing ρ_{KL} can be formulated as the optimization

$$\begin{aligned}\max_{u\in\mathcal{U}} \ & \min\{\gamma_1, \gamma_2\} \\ \text{s.t.} \ & \tfrac{1}{2}\big(\ln\det\Sigma_2 - \ln\det\Sigma_1 + \mathrm{Tr}\,\Sigma_2^{-1}\Sigma_1 \\ & \quad +(\mu_1-\mu_2)^{\mathrm{T}}\Sigma_2^{-1}(\mu_1-\mu_2) - n\big) \geq \gamma_1, \\ & \tfrac{1}{2}\big(\ln\det\Sigma_1 - \ln\det\Sigma_2 + \mathrm{Tr}\,\Sigma_1^{-1}\Sigma_2 \\ & \quad +(\mu_1-\mu_2)^{\mathrm{T}}\Sigma_1^{-1}(\mu_1-\mu_2) - n\big) \geq \gamma_2,\end{aligned} \tag{17}$$

where the mean and covariance are convex functions of the control input u. However, neither optimization (17) is not convex, even for the case when the expectations μ_1 and μ_2 are linearly dependent on $u \in \mathcal{U}$, and the covariances Σ_1 and Σ_2 are independent of $u \in \mathcal{U}$.

B. Using ρ_{geo}

Consider the measure of statistical distance ρ_{geo} between two Gaussian distributions. Maximizing ρ_{geo} can be formulated as

$$\begin{aligned}\min_{u\in\mathcal{U}} \max_{\mu,\gamma} \ & -\gamma \\ \text{s.t.} \ & \begin{bmatrix} \gamma & (\mu-\mu_1(u))^{\mathrm{T}} \\ (\mu-\mu_1(u)) & \Sigma_1(u) \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} \gamma & (\mu-\mu_2(u))^{\mathrm{T}} \\ (\mu-\mu_2(u)) & \Sigma_2(u) \end{bmatrix} \succeq 0.\end{aligned} \tag{18}$$

Lemma 1. Suppose that $\mu_i : \mathcal{U} \to \mathbb{R}^n$ is an affine function and $\Sigma_i : \mathcal{U} \to \mathbb{S}^n_+$ is an affine or concave quadratic function for each $i = 1, 2$. Then, the optimization (18) is a convex-constrained concave program, i.e., the objective function is concave and the constraint set is convex.

Since the optimization (18) has a concave objective function and the constraint set is convex and compact, a global optimum is achieved on the boundary of the closed convex constraint set, i.e., $u^\star \in \partial\mathcal{U}$ where $\partial\mathcal{U}$ refers to the boundary of $\mathcal{U}$, or must be constant over $\mathcal{U}$. An iteration algorithm is described in Algorithm 1. Furthermore, if the input constraint set $\mathcal{U}$ is a polytope then an global optimal solution $u^\star$ is in the set of vertices, denoted by $\mathcal{U}_{\text{v}}$. This implies that for a polytope $\mathcal{U}$, we only need to compare a finite number of candidates for optimal inputs:

$$u^\star = \arg\max_{u\in\mathcal{U}_{\text{v}}} \rho_{\text{geo}}(\mathcal{N}(\mu_1(u), \Sigma_1(u)), \mathcal{N}(\mu_2(u), \Sigma_2(u))).$$

C. One-step Maximization Detectability or Distinguishability of Two Competing Faults Using ρ_{geo}

Consider two competing fault models M_i $(i = 1, 2)$ in (5) and assume perfect information of the state variables[§] and that the output transition maps are dropped for simplicity (the extension to the general case is straightforward). Suppose that the current state of the true process has a known Gaussian distribution $\mathcal{N}(\bar{x}_t, \Sigma_{x_t})$, the dimensions of x^i are the same as the true state's, and $x_t = x_t^i$ for $i = 1, 2$. Due to linearity of the state transition map of models M_i, the one-step lookahead trajectories of the models M_i $(i = 1, 2)$ corresponding to two hypothesized faults are also Gaussian, provided an affine state feedback or open-loop control u_t. An optimal input design can be

[§]By *perfect information*, we mean that the state vector is measurable and its distribution is known or an unbiased estimation can be obtained with the associated computable error covariance.

formulated as an optimization of finding $u \in \mathcal{U}$ maximizing $\rho_{\text{geo}}(\mathcal{N}(\bar{x}^{f_1}_{t+1}, \Sigma_{x^{f_1}_{t+1}}), \mathcal{N}(\bar{x}^{f_2}_{t+1}, \Sigma_{x^{f_2}_{t+1}}))$. Consider an affine state feedback control

$$u_t = K_t x_t + \nu_t. \tag{20}$$

Then, the mean and covariance of the one-step lookahead state of the model M_i are

$$\begin{aligned} \bar{x}^i_{t+1} &= (A^i_t + B^i_t K_t)\bar{x}_t + B^i_t \nu_t \\ \Sigma_{x^i_{t+1}} &= (A^i_t + B^i_t K_t)\Sigma_{x_t}(A^i_t + B^i_t K_t)^{\mathrm{T}} + E^i_t \Sigma_w (E^i_t)^{\mathrm{T}} \end{aligned} \tag{21}$$

for each $i = 1, 2$. For input constraints, consider $\mathcal{U}_t \triangleq \{u \in \mathbb{R}^{n_u} : H_{u,t}\mathbf{E}[u] \le b_{u,t}$ and $\mathbf{Var}[u] \preceq \Sigma^{\max}_{u,t}\}$ where the subscript t denotes time dependence and $\Sigma^{\max}_{u,t}$ refers to an upper bound on the covariance of interest. For an affine state feedback control (20), $u_t \in \mathcal{U}_t$ if and only if K_t and ν_t satisfy the inequalities

$$\begin{aligned} H_{u,t}K_t\bar{x}_t + H_{u,t}\nu_t &\le b_{u,t} \\ \begin{bmatrix} \Sigma^{\max}_{u,t} & K_t \\ K_t^{\mathrm{T}} & \Sigma^{-1}_{x_t} \end{bmatrix} &\succeq 0 \end{aligned} \tag{22}$$

that are convex in (K_t, ν_t). The resulting optimization for an optimal input design with the input constraint (22) is given by (23).

In addition to input constraints, to avoid instability and performance degradation of the closed-loop system, the state constraints for x^i_{t+1} $(i = 1, 2)$ should also be considered. Since the predicted state trajectories are essentially stochastic, the corresponding state constraints are written in terms of chance constraints. Similar to input constraints, consider $\mathcal{X}_t \triangleq \{x \in \mathbb{R}^n : H_{x,t}\mathbf{E}[x] \le b_{x,t}$ and $\mathbf{Var}[x] \preceq \Sigma^{\max}_{x,t}\}$.

Algorithm 1 Iteration algorithm for solving (18).

Input: $\mu_i(\cdot), \Sigma_i(\cdot)$; $i = 1, 2$, $u^{(0)}$, and $\{\delta^{(j)}\} \subset \mathbb{R}_{++}$.
Onput: $\hat{u}^*, \hat{\gamma}^*$, and δ.

Step 0: Set $j = 0$.
Step 1: For $u := u^{(j)}$, solve the minimization part in (18) and assign its optimal value by $\gamma^{(j)} := \gamma^*$.
Step 2: Set $u := u^{(j)} + du$.
Step 3: Solve the minimization

$$\begin{aligned} \min_{u \in \mathcal{U}, \mu, \gamma} \quad & \gamma \\ \text{s.t.} \quad & \begin{bmatrix} \gamma & (\mu - \mu_1(u))^{\mathrm{T}} \\ (\mu - \mu_1(u)) & \Sigma_1(u) \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} \gamma & (\mu - \mu_2(u))^{\mathrm{T}} \\ (\mu - \mu_2(u)) & \Sigma_2(u) \end{bmatrix} \succeq 0, \\ & \gamma \ge (1 + \delta^{(j)})\gamma^{(j)}. \end{aligned} \tag{19}$$

if (19) has a feasible optimal solution (γ^*, du^*, μ^*) **then**
Assign the optimum and optimal value by $u^{(j+1)} := u^{(j)} + du^*$ and $\gamma^{(j+1)} := \gamma^*$. Set $j := j + 1$ and go to Step 2.
else
Set $\hat{u}^* := u^{(j)}$, $\hat{\gamma}^* := \gamma^{(j)}$, and $\delta := \delta^{(j)}$.
end if

From (21), $x^i_{t+1} \in \mathcal{X}_t$ if and only if

$$\begin{aligned} H_{x,t}(A^i_t + B^i_t K_t)\bar{x}_t + H_{x,t}B^i_t\nu_t &\le b_{x,t} \\ \begin{bmatrix} \Sigma^{\max}_{x,t} - E^i_t\Sigma_w(E^i_t)^{\mathrm{T}} & (A^i_t + B^i_t K_t) \\ (A^i_t + B^i_t K_t)^{\mathrm{T}} & \Sigma^{-1}_{x_t} \end{bmatrix} &\succeq 0. \end{aligned} \tag{25}$$

that are convex in (K_t, ν_t). The resulting optimization for an optimal input design with the input and state constraints (22) and (25) is given by (24).

Remark 3. The convex constraints (22) and (25) are the intersections of a polytope and a positive-semidefinite cone (basically, a mixed linear-conic constraint).

Algorithm 1 cannot be directly applied to solve the optimization problems (23) and (24) for the decision variable (K_t, ν_t). However, if a state feedback gain K_t is fixed then Algorithm 1 can be used to solve those optimizations for ν_t. Computing a state-feedback gain K_t in the optimizations (23) and (24) can be performed separately by solving the LMIs in (22) and (25), respectively. Note that solutions of such convex constraints are not generally unique. To resolve this non-uniqueness of feasible solutions K_t, the associated symmetric matrices defining the LMIs could be forced to be close to the extreme rays of positive-semidefinite cone, which is the set of rank-one symmetric matrices. This could be achieved by minimizing the rank of the resulting symmetric matrices and there are several ways to approximately perform rank-minimization using smooth approximation to the rank operator for positive-semidefinite matrices. For example, $-\log\det(X)$, which is convex in $X \in \mathbb{S}^n_{++}$ (or $X \in \mathbb{S}^n_+$),[¶] or Tr(X) which is linear can be used.

D. A Separate Design Method of State Feedback using $\mathcal{H}_2$ Optimal Control

Consider the fault scenario models (5) for $i = 1, 2$, where the measurement noise v^i_t is ignored without loss of generality.[‖] Since the $\mathcal{H}_2$ norm can be interpreted as the maximum output variance excited by white noise of the unit $\mathcal{L}_2$ norm, a natural way to compute a state feedback control gain K_t for a fixed $t \ge 0$ satisfying the constraints on the variance is $\mathcal{H}_2$-optimal control [18]. Utilizing the LMI conditions for state feedback $\mathcal{H}_2$ synthesis,[**] for each time instance $t \ge 0$, design of a state feedback gain K_t can be performed by the following procedure.

S1. Solve the following SDP for Z, X, and W: Minimize η subject to

$$\begin{aligned} \begin{bmatrix} A^i_t & B^i_t \end{bmatrix} \begin{bmatrix} X \\ Z \end{bmatrix} + \begin{bmatrix} X & Z^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} A^{i\,\mathrm{T}}_t \\ B^{i\,\mathrm{T}}_t \end{bmatrix} + E^i_t E^{i\,\mathrm{T}}_t &\prec 0, \\ \begin{bmatrix} X & (C^i_t + D^i_t Z)^{\mathrm{T}} \\ (C^i_t + D^i_t Z) & W \end{bmatrix} &\succ 0, \\ \mathrm{Tr}(W) &< \eta, \\ X &\succ 0. \end{aligned} \tag{26}$$

Define the optimal solutions as $Z_\star$, $X_\star$, and $W_\star$.

¶ If we consider $\mathbb{S}^n_+$ as the domain of the function $-\log\det(\cdot)$ then an extended real line is considered as the range of $-\log\det(\cdot)$, i.e., $-\log\det : \mathbb{S}^n_+ \to (-\infty, \infty]$.

‖ The perturbation or variation due to v^i_t is not controllable by using a state feedback controller.

** See [18] for details on $\mathcal{H}_2$-synthesis problems.

$$\max_{K_t,\nu_t} \min_{\mu,\gamma} \gamma$$
$$\text{s.t.} \begin{bmatrix} \gamma & (\mu-(A_t^i+B_t^iK_t)\bar{x}_t+B_t^i\nu_t)^{\mathrm{T}} \\ (\mu-(A_t^i+B_t^iK_t)\bar{x}_t+B_t^i\nu_t) & (A_t^i+B_t^iK_t)\Sigma_{x_t}(A_t^i+B_t^iK_t)^{\mathrm{T}}+E_t^i\Sigma_w(E_t^i)^{\mathrm{T}} \end{bmatrix} \succeq 0,\ i=1,2,$$
$$\begin{bmatrix} \Sigma_{u,t}^{\max} & K_t \\ K_t^{\mathrm{T}} & \Sigma_{x_t}^{-1} \end{bmatrix} \succeq 0,\ b_{u,t}-H_{u,t}K_t\bar{x}_t-H_{u,t}\nu_t \geq 0. \tag{23}$$

$$\max_{K_t,\nu_t} \min_{\mu,\gamma} \gamma$$
$$\text{s.t.} \begin{bmatrix} \gamma & (\mu-(A_t^i+B_t^iK_t)\bar{x}_t+B_t^i\nu_t)^{\mathrm{T}} \\ (\mu-(A_t^i+B_t^iK_t)\bar{x}_t+B_t^i\nu_t) & (A_t^i+B_t^iK_t)\Sigma_{x_t}(A_t^i+B_t^iK_t)^{\mathrm{T}}+E_t^i\Sigma_w(E_t^i)^{\mathrm{T}} \end{bmatrix} \succeq 0,\ i=1,2,$$
$$\begin{bmatrix} \Sigma_{x,t}^{\max}-E_t^i\Sigma_w(E_t^i)^{\mathrm{T}} & (A_t^i+B_t^iK_t) \\ (A_t^i+B_t^iK_t)^{\mathrm{T}} & \Sigma_{x_t}^{-1} \end{bmatrix} \succeq 0,\ b_{x,t}-H_{x,t}(A_t^i+B_t^iK_t)\bar{x}_t-H_{x,t}B_t^i\nu_t \geq 0,\ i=1,2,$$
$$\begin{bmatrix} \Sigma_{u,t}^{\max} & K_t \\ K_t^{\mathrm{T}} & \Sigma_{x_t}^{-1} \end{bmatrix} \succeq 0,\ b_{u,t}-H_{u,t}K_t\bar{x}_t-H_{u,t}\nu_t \geq 0. \tag{24}$$

S2. Solve the following SDP for Z and W: Minimize η subject to

$$\begin{aligned} \begin{bmatrix} A_t^i & B_t^i \end{bmatrix} \begin{bmatrix} X_\star \\ Z \end{bmatrix} + \begin{bmatrix} X_\star & Z^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} A_t^{i\,\mathrm{T}} \\ B_t^{i\,\mathrm{T}} \end{bmatrix} + E_t^i E_t^{i\,\mathrm{T}} &\prec 0, \\ \begin{bmatrix} X_\star & (C_t^i+D_t^iZ)^{\mathrm{T}} \\ (C_t^i+D_t^iZ) & W \end{bmatrix} &\succ 0, \\ \mathrm{Tr}(W) &< \eta, \\ \begin{bmatrix} \Sigma_{u,t}^{\max} & Z \\ Z^{\mathrm{T}} & X_\star\Sigma_{x_t}^{-1}X_\star \end{bmatrix} &\succeq 0, \\ \begin{bmatrix} \Sigma_{x,t}^{\max}-E_t^i\Sigma_w(E_t^i)^{\mathrm{T}} & (A_t^iX_\star+B_t^iZ) \\ (A_t^iX_\star+B_t^iZ)^{\mathrm{T}} & X_\star\Sigma_{x_t}^{-1}X_\star \end{bmatrix} &\succeq 0, \end{aligned} \tag{27}$$

for $i=1,2$. Define the optimal solutions as $Z^\star$ and $W^\star$.

S3. Compute the state feedback control gain $K_t = Z^\star X_\star^{-1}$.

Once the state feedback gain K_t is fixed, the remaining problem is to determine the affine term ν_t solving the constrained optimization (23) or (24), which can be performed, again, by using Algorithm 1.

VI. Conclusion

This paper considers optimal *active* input design problems for fault detection and diagnosis based on Bayesian inference. The resulting optimization for input design is to maximize statistical discrimination between models of hypotheses corresponding to fault scenarios. Each model of a fault scenario characterizes a random process of the measurable outputs and, to quantify the quality of the measurable data for FDD, this paper considers the well-known KL-divergence and its approximation using geometric properties of confidence ellipsoids associated with Gaussian random processes. With such statistical distance measures, the original optimization is non-convex even without any constraints, which can be computationally intractable, especially for multiple hypothesis tests. We propose a sequential SDP method to find a local optimum that can be further improved by using multiple shooting or warm starts. In addition, closed-loop state feedback input design problems are proposed, for which semi-chance constraints are introduced to impose bounds on the expected controlled trajectories and their variances.

Acknowledgement

The authors would like to acknowledge Rolf Findeisen at Otto-von-Guericke University Magdeburg for reading helpful suggestions and comments, and Lixian Zhang at Harbin Institute of Technology for valuable discussions.

References

[1] T. Henzinger, "The theory of hybrid automata," in *Proc. of LICS*, 1996, pp. 278–292.

[2] M. W. Hofbaur, *Hybrid Estimation of Complex Systems*. Berlin: Springer-Verlag, 2005.

[3] 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.

[4] 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.

[5] S. L. Campbell and R. Nikoukhah, *Auxiliary Signal Design for Failure Fetection*. Princeton, NJ: Princeton University Press, 2004.

[6] H. H. Niemann, "A setup for active fault diagnosis," *IEEE Trans. on Auto. Contr.*, vol. 51, pp. 1572–1578, 2006.

[7] M. Simandl and I. Puncochar, "Active fault detection and control: unified formulation and optimal design," *Automatica*, vol. 45, pp. 2052–2059, 2009.

[8] J. K. Scott, R. Findeisen, R. D. Braatz, and D. M. Raimondo, "Design of active inputs for set-based fault diagnosis," in *In Press: Proc. of the 2013 American Control Conference*, 2013.

[9] 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.

[10] H. H. Niemann and J. Stoustrup, "Robust fault detection in open loop vs. closed loop," in *IEEE CDC*, 1997, pp. 4496–4497.

[11] 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.

[12] M. Simandl, I. Puncochar, and J. Kralovec, "Rolling horizon for active fault detection," in *IEEE CDC*, 2005, pp. 3789–3794.

[13] A. E. Ashari, R. Nikoukhah, and S. Campbell, "Effects of feedback on active fault detection," *Automatica*, vol. 48, pp. 866–872, 2012.

[14] ——, "Active robust fault detection in closed-loop systems: Qquadratic optimization approach," *IEEE Trans. on Auto. Contr.*, vol. 54, pp. –, 2012, in press.

[15] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. UK: Cambridge University Press, 2003.

[16] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Motivations to Computational Implementation*. NY, USA: Springer-Verlag, 2001.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: John Wiely & Sons, Inc., 2006.

[18] G. E. Dullerud and F. Paganini, *A Course in Robust Control Theory: A Convex Approach*. New York: Springer, 2000.