Fault Isolation in Data Driven Multivariate Process Monitoring

Dimitry Gorinevsky*, *Fellow IEEE*

***Abstract*—Consider a training set of multivariate input/output process data. Given a new observation we ask the following questions: is the new observation normal or abnormal? is one of the inputs or outputs abnormal (faulty) and which? For a linear Gaussian model of the process, the problem is solved by Bayesian hypothesis testing. The formulation differs from existing multivariate statistical monitoring methods by taking variance (uncertainty) of the linear regression model into account. In the limit case of zero model variance, the proposed method matches the established methods for anomaly detection and fault isolation. The proposed method might yield an order of magnitude reduction in fault isolation errors compared to the established approaches when regression models have large variance. This is the case for ill-conditioned multivariate regression models even with large training data sets. The paper also shows that isolating faults to a small ambiguity group works much better than trying to isolate a single fault. The proposed method is verified in a Monte Carlo study and in application to jet engine fault isolation.**

I. INTRODUCTION

This paper considers a problem of multivariate statistical process monitoring using data driven models. This section introduces the problem informally.

A. Problem

Consider a data set consisting of a series of tuples

$$\mathcal{D}_N = \{x(t), y(t)\}_{t=1}^{N}, \tag{1}$$

where t is the time series index, $x(t) \in \Re^n$, and $y(t) \in \Re^m$. Data (1) is used for regression modeling, where $x(t)$ are the independent variables (process inputs, regressors, explanatory variables) and $y(t)$ are dependent variables (process outputs, response variables, quality variables). The inputs $x(t)$ are observable or known. The outputs $y(t)$ are generated by a (nominal) random process with $x(t)$ as inputs. The nominal random process is stationary and observations $\{x(t), y(t)\}$, $\{x(s), y(s)\}$ are independent for $s \neq t$.

Given $\mathcal{D}_N$ and a new observation $\{x, y\}$ we ask the following questions: is the new observation normal or abnormal? is one of the observed inputs x_j or outputs y_k biased (faulty) and which? This paper considers an easy generalization of the above questions. We assume that the input and output faults have signatures defined by the two sets

$$F = \{f_1, ..., f_n\}, \tag{2}$$

$$G = \{g_1, ..., g_m\} \tag{3}$$

As a special case, a fault in a single input channel j is described by $f_j = e_j$, where $e_j \in \Re^n$ is a unit vector. A fault in an output channel k is described by $g_k = e_k$, where $e_k \in \Re^m$ is a unit vector. The following statistical hypotheses are evaluated to answer the questions.

$\mathbf{H}_N$: Null hypothesis. Observation $\{x, y\}$ is generated by the nominal process.

$\mathbf{H}_A$: Anomaly. Observation $\{x, y\}$ is abnormal, deviates from the nominal process. This is modeled as observing the output $y + h$ instead of y, where h is a nuisance vector.

$\mathbf{H}_{I,j}$: Input channel fault. An anomaly, where y is generated by the nominal process with x. The input $x + zf_j$ is observed instead of x, where z is a scalar nuisance parameter.

$\mathbf{H}_{O,k}$: Output channel fault. An anomaly where y is generated by the nominal process with x. The output $y + zg_k$ is observed instead of y; z is a scalar nuisance parameter.

We consider the following anomaly detection and fault isolation problems

ANO: Determine which of the two complementary hypotheses holds: $\mathbf{H}_N$ nominal or $\mathbf{H}_A$ anomaly.

ISO: If $\mathbf{H}_A$ holds, find a set of all likely fault hypotheses $\mathbf{H}_{I,j}$ and $\mathbf{H}_{O,k}$ (the ambiguity group).

MAP: Find the most likely hypothesis out of the hypotheses: $\mathbf{H}_N$, $\mathbf{H}_A$, $\mathbf{H}_{I,j}$, and $\mathbf{H}_{O,k}$.

Problems ANO and MAP have been considered in the prior work. Problem ISO of finding the fault ambiguity group appears to be new. If the anomaly is detected but the ambiguity group is empty, we report "an unknown anomaly"; a related hypothesis is considered in [2]. The fault ambiguity groups are considered in the literature on circuit testing and discrete fault diagnostics from discrete data, but not for the stated problem.

B. Motivation

The proposed data-driven formulation is aimed at isolating faults that induce bias in one of the raw data channels. The data $\{x, y\}$ might be combinations of the raw data channels.

The assumption of $x(t)$ being observable and $y(t)$ a random variable realization is standard in regression modeling. The formulation with jointly normally distributed $x(t)$, $y(t)$ is, in fact, a special case of (1): the input is empty and the output is a stacked vector of $x(t)$, $y(t)$.

This work was supported by NASA Grant NNX07AEIIA, NASA Contract NNX12CA02C, and Air Force Contract FA8650-10-C-2006

Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305, gorin@stanford.edu and Mitek Analytics LLC, Palo Alto, CA 94306, dimitry@mitekan.com

It is implicitly assumed that the anomalies have small probability. This is often true in monitoring of physical systems and is the reason for using training data (1) without anomalies. The statistical hypotheses with one fault only are considered. Small anomaly probability makes unlikely that two different faults would occur simultaneously.

C. Baseline approach

An established approach to solving the stated monitoring problem is to fit a linear regression model to the training data (1) and then use this model for statistical testing of the hypotheses. We recite it in this section and later use as a baseline for introducing the proposed approach. Some of the baseline method literature is surveyed in Section II below.

In what follows, it is assumed that process output $y(t)$ data are realizations of independent random variables y that are normally distributed, conditional on $x(t)$

$$y(t) \sim N(Bx(t), S). \tag{4}$$

Model matrix $B \in \Re^{m,n}$ and covariance matrix $S \in \Re^{n,n}$ define the regression model. To estimate B and S, the training data in $\mathcal{D}_N$ can be formed into two data matrices

$$X = [x(1), ..., x(N)] \in \Re^{n,N}, \tag{5}$$
$$Y = [y(1), ..., y(N)] \in \Re^{m,N}. \tag{6}$$

The least squares estimates of the parameters of regression model (4) are well known to be

$$B_N = YX^T \left(XX^T\right)^{-1}, \tag{7}$$
$$S_N = N^{-1}(Y - B_N X)(Y - B_N X)^T. \tag{8}$$

For now, we assume that matrices XX^T and VV^T, where $V = Y - B_N X$, are invertible. The case of non-invertible XX^T and VV^T is discussed in Subsection III-C

Assuming that the estimates (7), (8) are accurate, a fault with signature h can be monitored through the following indexes

$$M_1(x, y) = \|y - B_N x\|^2_{S_N^{-1}}, \tag{9}$$
$$M_2(x, y; h) = \min_z \|y - B_N x - zh\|^2_{S_N^{-1}}, \tag{10}$$
$$z = (y - B_N x)^T S_N^{-1} h / \|h\|^2_{S_N^{-1}}, \tag{11}$$

where notation $\|h\|^2_{S_N^{-1}} = h^T S_N^{-1} h$ is used. Index M_1 indicates an anomaly when it exceeds a threshold. Index M_2 is used to perform fault isolation; it allows to detect the fault with the signature h. The optimal fault amplitude estimate z in (10) is given by (11). For the input fault in channel j, the assumed signature is $h = B_N f_j$; for the output fault k, the signature is $h = g_k$. The fault hypothesis providing the smallest index M_2 is assumed to hold.

D. Model variance in anomaly detection

Anomalies can be detected using (9) assuming that model (4) is perfectly estimated, $B = B_N$, $S = S_N$. Under this assumption, index M_1 (9) follows the χ^2_m distribution. Accordingly, the threshold in detecting anomaly through M_1 can be chosen as the p-value for the χ^2_m distribution to provide an acceptable False Positive (FP) rate.

In fact, the estimates B_N, S_N obtained from a finite sample contain an error. The variance of B_N, S_N has been earlier taken into in account in the special case of $x(t) = 1$, $n = 1$ in (1). In that case, B is the mean of the normal distribution (4) and B_N is the estimate of the mean. In [15], it was established that index M_1 (9) then follows the Hotelling's T^2 distribution, and the p-value for the T^2 should be used instead of the χ^2_m.

This paper addresses the impact of the variance of B_N, S_N in the general case of the regression model (4). This general case does not appear to be addressed in the existing literature. The only related work seems to be detection of an unknown signal in a noisy channel discussed in [23], which includes results related to anomaly detection herein.

If XX^T is ill-conditioned, the variance of B_N can be large even for large N. As a result, the decisions based on (9)–(10) may be far from optimal. Issues with accuracy of estimation of large covariance matrices from the data have recently attracted attention in the signal processing domain, see [6].

E. Fault isolation with model uncertainty

Most of the fault isolation work assumes the model is either given or accurately estimated from the data. There are few exceptions, however, where model variance is considered.

In [31], confidence intervals, which are related to the model variance, are used to test significance of linear combinations of regression coefficients. The results of [31] are related to intermediate steps of the proposed solution.

Optimal Bayesian estimation taking into account gaussian uncertainty of linear models is considered in [30], [43]. These are nonconvex problems. The fault isolation problems in this paper are easier to solve because just a single scalar fault intensity parameter is estimated for each fault hypothesis.

F. Contributions

The main novelty of this work is that indexes (9), (10) are modified to explicitly address variance (uncertainty) of the model trained on finite amount of data.

Subsection I-A states Problem ISO of finding the fault ambiguity group. This problem appears to be new. The paper demonstrates that isolating the single channel faults to the ambiguity group can make the isolation errors uniformly small.

One more contribution of this paper is in establishing the tuning rules that are based on asymptotic performance of the algorithms and allow to set up the algorithm parameters.

The simulation results below show that the proposed Bayesian approach can reduce fault isolation errors significantly, in some cases, by an order of magnitude.

G. Outline

The paper outline is as follows. Section II surveys the related work. Section III introduces the optimal Bayesian log-posteriors for the hypotheses. Section IV describes the proposed algorithms based on hypothesis testing. Section V considers asymptotic performance of the algorithms and tuning rules based on this performance. Finally, Sections VI and VII illustrate the proposed method with numerical examples.

II. Related prior work

Section I described the main contribution of this paper - improving on the baseline approach by addressing the variation of the estimated model. There is also substantial literature on the baseline approach, other methods for anomaly detection and fault isolation, and their applications. This section surveys such work, which is less directly related to the paper contribution.

A. Multivariate Statistical Process Control

The formulated problems are related to Multivariate Statistical Process Control (MSPC) methodology of statistical process monitoring, developed for process control applications and surveyed in [1], [24], [29], [34]. The MSPC approaches generally follow the baseline approach in Section I. The matrices B and S in the gaussian model (4) are estimated from the data. The estimates are then used for the anomaly detection and fault isolation assuming they are accurate.

Early MSPC work on fault isolation was based on geometrical argument of subspace projections, distances, and angles for the monitored data, e.g., see [34]. More recent paper [1] argues that an index similar to (10), (11) with added regularization, the fault reconstruction index, provides the best results. The closest related work uses Bayesian formulations for testing fault hypotheses. The tests in [18] are similar to (11), but restricted to the null space of the covariance matrix. This paper uses the same fundamental approach as [21], [28]: the fault isolation is done only after the null hypothesis is rejected.

B. Covariance structure

Decision indexes (9), (10) imply that the empirical covariance S_N is invertible. Subsection III-C discusses how an non-invertible covariance S_N can be handled by introducing Bayesian priors (regularization).

The problem of non-invertible S_N is well-studied in the special case of $x(t) \in \varnothing$ (there are no regressors). In this case, $S_N = YY^T$ in (8). The Principal Component Analysis (PCA) approaches decompose the rank-deficient covariance matrix S_N into the null space and the range space operators. The range space projection of the data can be addressed by the baseline approach; in this space the covariance operator is invertible. The decision index (9) for the range space projection is usually called T^2 statistics (this name might be not quite precise). The null space projection is described by Q statistics, a.k.a., the square prediction error (SPE).

The anomaly detection might use both T^2 and Q, such as in [3]. Alternatively, [18] considers the range space variation as a nuisance variable and monitors the null space only.

Fault isolation using a single "reconstruction index" combining T^2 and Q and similar to (10), (11) is suggested in [1]. Effectively, [1] replaces the singular values σ_k of the matrix S_N with $\hat{\sigma}_k = \max(\sigma_k, \mu)$, where μ is a small regularization parameter. In Subsection III-C below, σ_k is replaced with $\hat{\sigma}_k = \sigma_k + \mu$. For most singular values, we have either $\sigma_k \gg \mu$ or $\sigma_k \ll \mu$ and the two approaches give very close results.

As mentioned in Section I-B, this paper considers $x(t)$ as a known input and $y(t)$ as a random variable output. One could consider $x(t)$ as a random variable as well. For example, Errors-in-Variables regression, see [40], assumes that $x(t)$ is measured with an random error. In that case, the variable transformation $[x(t)^T\, y(t)^T]^T \to y(t)$ and $\varnothing \to x(t)$, yields a special case of the problem considered in this paper. The connection with the Errors-in-Variables results is discussed in the end of Subsection III-A. Canonical-correlation analysis (CCA) and Partial Least Squares (PLS) [26] are two other approaches that consider $x(t)$ as a random vector and analyze the covariance of $[x(t)^T\, y(t)^T]^T$.

C. Model-based fault diagnostics

The control systems literature includes much work in model-based fault detection and isolation (FDI), e.g., see [4], [10]. This work assumes the system model as given; this is fundamentally different from the data-driven modeling in this paper. Apart from that, the linear dynamic models used in the FDI work are equivalent to the static models in this paper. They can be transformed into form (4) by applying a lifting transformation and noise whitening, e.g., see [4].

Some of the model-based FDI work considers noiseless data, assuming known model B in (4) and zero covariance, $S = 0$, e.g., see [4], [10]. The robust model-based FDI work describes model uncertainty as the process noise, see [4]. The robust model-based FDI methods require knowledge of all model parameters, including the model uncertainty, at the outset. The approach of this paper computes the uncertainty (the variation of the estimated model) along with the model, from the data. Such data-driven approach is easier to apply in practice.

Some of the model-based fault detection and isolation work uses Bayesian formulations for testing fault hypotheses. For example, using filter banks for testing alternative fault hypothesisis is discussed in [22], [36], [38], [45]. These approaches are, in spirit, related to the hypothesis testing for fault isolation in this paper, though the exact formulations are different.

D. Limitations of proposed approach

This paper assumes that no more than one fault at a time can occur. If multiple faults are possible, the fault isolation problem has combinatorial complexity, e.g., see [21]. This problem can be handled though ℓ_1 relaxation, e.g., see [14]. The ℓ_1 relaxation of the problem with the inputs x and outputs y including a mixture of discrete (binary) and real components is presented in [46].

The approach in this paper assumes linear Gaussian model (4). Such models cover a broad range of practical applications. This paper does not address the use of nonlinear models in FDI, and non-gaussian noises. One more important extension not adressed herein is anomaly detection and fault isolation for a population of systems in a multi-level regression framework, see [5], [12].

III. Bayesian formulation

This section formulates Bayesian framework for testing the ANO, ISO, and MAP hypotheses of Section I taking into

account the model uncertainty. The hypothesis likelihoods are computed assuming that data (5), (6) follow model (4). These likelihoods are next used in Section IV to solve the anomaly detection and fault isolation problems introduced in Section I. Subsection III-A discusses the likelihood for the null hypothesis. Subsection III-B describes the likelihoods for the fault hypotheses. Subsection III-C introduces the priors.

A. Log-likelihood index

The probability density for $y \sim N(Bx, S)$ in (4) yields the single-point log-likelihood

$$\begin{aligned} L(y, B, S|x) &= \log p(y, B, S|x) + c \\ &= -\frac{1}{2}(y - Bx)S^{-1}(y - Bx) - \frac{1}{2}\log\det(S), \end{aligned} \tag{12}$$

where $c = \frac{m}{2}\log(2\pi)$. Since all data samples are independent, the Maximum Likelihood Estimation (MLE) of B and S from i.i.d. data $\mathcal{D}_N$ (1) is obtained by solving the problem

$$L_N = \max_{B,S} \sum_{t=1}^{N} L(y(t), B, S|x(t)). \tag{13}$$

Solution of multivariate linear regression (12), (13) is well known. It is briefly recapped as a part of the derivation in Appendix A: optimal B, S are given by (7), (8); the optimal index (13) is

$$L_N = -\frac{N}{2}\log\det\left[\frac{Y\left[I - X^T(XX^T)^{-1}X\right]Y^T}{N}\right] - \frac{Nm}{2}. \tag{14}$$

One possible approach would be to compute the uncertainty (variation) of matrices B, S estimated from the training data $\mathcal{D}_N$. The uncertainty could then be used in the formulation for the new observation $\{x, y\}$ (the test set). This paper shows that it is much more convenient to consider the joint Bayesian formulation for all observations instead. By considering the new observation $\{x, y\}$ to be a part of the training data (1), we get the extended log-posterior index

$$L_+ = \max_{B,S}\left(L(y, B, S|x) + \sum_{t=1}^{N} L(y(t), B, S|x(t))\right). \tag{15}$$

Appendix A expresses the solution of (15) through the solution of (13) in the form

$$L_+(y|x) = c_N - \frac{N+1}{2}\log\left(1 + \frac{M_+(x,y)}{N}\right), \tag{16}$$

$$M_+(x,y) = \frac{\|y - B_N x\|^2_{S_N^{-1}}}{1 + N^{-1}\|x\|^2_{Q_N^{-1}}}, \tag{17}$$

where S_N (8), L_N (14) are computed from training data (5), (6), along with B_N, c_N, and Q_N

$$\begin{aligned} c_N &= \frac{(N+1)m}{2}\log(1 + N^{-1}) - \frac{N+1}{N}L_N, \\ Q_N &= N^{-1}XX^T, \quad B_N = N^{-1}YX^TQ_N^{-1}. \end{aligned} \tag{18}$$

In (16), $L_+(y|x)$ is a monotonically decreasing function of $M_+(x, y)$. The problem of maximizing $L_+(y|x)$ thus can be solved by minimizing $M_+(x, y)$. Note, that the denominator in (17) manifests *regression dilution* effect, also known as *attenuation*. The effect is well discussed in connection to errors-in-variables statistical models. The statistical uncertainty within the errors-in-variable framework could be introduced by considering predictor variable x normally distributed and Q_N its empirical covariance.

B. Hypothesis likelihoods

The posteriors for the hypotheses introduced in Section I for fault signatures (2), (3) can be expressed using (16)–(18).

We assume that the hypotheses priors are

$$\begin{aligned} P(\mathbf{H_N}) = 1 - p_A, \;\; P(\mathbf{H_A}) = p_A, \\ P(\mathbf{H_{I,j}}) = p_F, \;\; P(\mathbf{H_{O,k}}) = p_F, \end{aligned} \tag{19}$$

where $\mathbf{H_A}$ and $\mathbf{H_N}$ are complementary hypotheses. All fault hypotheses are assumed to have the same prior probabilities p_F. The generalization to different fault priors is straightforward, but would reduce clarity. The posterior log-likelihood of each hypothesis can be computed as

$$L(\mathbf{H}) = \log\left[P(x, y|\mathbf{H}) \cdot P(\mathbf{H})\right] = L_+(y|x, \mathbf{H}) + \log p_H, \tag{20}$$

where $\mathbf{H}$ is one of hypotheses in (19) and p_H is the respective probability. We used the Bayes rule: $P(\mathbf{H}|x, y) = P(x, y|\mathbf{H})P(\mathbf{H})$const, then $L_+(x, y|\mathbf{H}) = \log P(x, y|\mathbf{H}) - \log$ const. The additive constant in (20) is ignored. The log posteriors for different hypotheses $\mathbf{H}$ are derived below.

$\mathbf{H}_N$*: Null hypothesis:* It is assumed that the observation $\{x, y\}$ is generated by the nominal process. Hence the derivation of Subsection III-A holds without any modifications. From (16), (17), (19), and (20), we get

$$\begin{aligned} L(\mathbf{H_N}) = c_N - \frac{N+1}{2}\log\left(1 + \frac{M_+(x,y)}{N}\right) \\ + \log(1 - p_A) \end{aligned} \tag{21}$$

$\mathbf{H}_A$*: Anomaly:* The anomaly hypothesis assumes that $y + h$ is observed instead of y, where h is a nuisance vector. From (16), (19) and (20) we get

$$\begin{aligned} L(\mathbf{H_A}) = c_N - \min_h \frac{N+1}{2}\log\left(1 + \frac{M_+(x, y+h)}{N}\right) \\ + \log p_A = c_N + \log p_A, \end{aligned} \tag{22}$$

In accordance with (17), $\min_h M_+(x, y + h) = 0$ is achieved for $h = -y + B_N x$.

$\mathbf{H}_{I,j}$*: Input channel fault:* The hypothesis of input fault j assumes that for the process data $\{x, y\}$, the observed input is $x + f_j z$ instead of x. Fault intensity z is unknown and input fault signature $f = f_j \in F$ is known.

From (16), (19), and (20) we get

$$\begin{aligned} L(\mathbf{H_{I,j}}) = c_N - \frac{N+1}{2}\log\left(1 + \frac{M_+(x + f_j z_{I,j}, y)}{N}\right) \\ + \log p_F, \end{aligned} \tag{23}$$

where $z_{I,j}$ is the most likely value of the nuisance parameter z. In accordance with (17),

$$z_{I,j} = \arg\min_z M_+[z], \tag{24}$$

$$M_+[z] = \frac{(y - B_N x - B_N f_j z)^T S_N^{-1}(y - B_N x - B_N f_j z)}{1 + N^{-1}(x + f_j z)^T Q_N^{-1}(x + f_j z)}.$$

The function minimized in (24) is the ratio of two quadratic polynomials in z, $M_+[z] = (az^2 + bz + c)/(dz^2 + ez + f)$. The derivative $dM_+[z]/dz$ has numerator that is quadratic in z. The minimum can be found by checking the two roots of this numerator.

$\mathbf{H}_{O,k}$: *Output channel fault:* The hypothesis of output fault k assumes that for the process data $\{x, y\}$, the observed output is $y+g_k z$ instead of y. Fault intensity z is unknown and output fault signature $g = f_j \in G$ is known.

From (16), (19), and (20), we get

$$L(\mathbf{H_{O,k}}) = c_N - \frac{N+1}{2}\log\left(1 + \frac{M_+(x, y + g_k z_{O,k})}{N}\right) + \log p_F, \quad (25)$$

where $z_{O,k}$ is the most likely value of the nuisance parameter z. Using (17) and differentiating to find minimum in z yields

$$\begin{aligned} z_{O,k} &= \arg\min_z M_+(x, y + g_k z_*) \\ &= (y - B_N x)^T S_N^{-1} g_k / \|g_k\|^2_{S_N^{-1}}. \end{aligned} \quad (26)$$

C. Priors for model parameters B and S

Non-invertible matrices XX^T and VV^T, in (7), (8), (17) can be handled through Bayesian priors that implicitly introduce regularization, shrinkage. Following the Bayesian approach, the MLE formulation (13) shall be replaced by Maximum A posteriori Probability (MAP) formulation including the priors for the matrices B and S in the distribution (4).

Consider a Normal-Inverse Wishart prior $P(B, S) = P(B|S)P(S)$, where $B|S$ follows a normal matrix distribution and S follows an Inverse Wishart distribution [19].

$$P(S) \propto (\det S)^{-\frac{1}{2}(p+m+1)} \exp\left(-\frac{\rho}{2}\text{trace}\, S^{-1}\right), \quad (27)$$

$$P(B|S) \propto \mu^{\frac{m}{2}} (\det S)^{-\frac{1}{2}m} \exp\left(-\frac{\mu}{2}\text{trace}\,(B^T S^{-1} B)\right). \quad (28)$$

The hyper-parameters μ and ρ describe the precisions (inverse covariances) for B and S respectively and p is a positive integer (the number of degrees of freedom). One can consider μ and ρ as small regularization parameters. In what follows, it is assumed that in the Inverse Wishart distribution $p \geq m+1$. In the numerical examples below, $p = m + 1$.

The MAP formulation adds a log-prior term $L_0 = \log P(B, S) = \log P(B|S) + \log P(S)$ following (27), (28), to the loss index in (13). The MLE formulation could be considered as a MAP with a noninformative prior. The Normal-Inverse Wishart prior is a conjugate prior of distribution (4) with unknown mean and covariance. This means the prior can be described as the posterior obtained with $p + 1$ pseudo-observations. The pseudo-observations are the columns of the matrices X_0 and Y_0, such that $X_0 X_0^T = \rho I$, $Y_0 X_0^T = 0$, $Y_0 Y_0^T = \mu I$.

With the introduced prior, the derivations presented above in the paper have the same form with the pseudo-observations added to the data (5), (6). In particular, matrices B_N, Q_N, (18), and S_N, (8), in the algorithms are replaced by

$$\begin{aligned} Q_N &= (p + N + 1)^{-1}(XX^T + \rho I), & (29)\\ B_N &= (p + N + 1)^{-1} Y X^T Q_N^{-1}, & (30)\\ S_N &= (p + N + 1)^{-1} & (31)\\ &\quad \times[(Y - B_N X)(Y - B_N X)^T + \mu I + \rho B_N B_N^T]. \end{aligned}$$

This ensures that the matrices S_N and Q_N are invertible. Note that the relative weight of the regularization parameters ρ and μ in (29), (30), (31) compared to the scatter matrix XX^T decreases as the number of the data points N increases making the regularization less effective. This should not be a problem if XX^T has a full rank asymptotically, which is the case for independent regressors. If this is not the case, one could drop the dependent regressors in an initial regressor set.

IV. Anomaly detection and fault isolation

This section presents the main contribution of the paper: the algorithms that solve the anomaly detection and fault isolation problems introduced in Section I. These Bayesian algorithms use the hypothesis posteriors computed in Section III. The problems are solved by computing and comparing log-posteriors (21), (22), (23), (25).

A. Monitoring for anomalies and the faults

Three monitors are described below. Each monitor is a computational function that inputs a data point $\{x, y\}$ and produces a list of one or more accepted hypotheses.

Monitor 1 (ANO): Anomaly detection: To determine if $\mathbf{H_N}$ nominal or $\mathbf{H_A}$ anomaly holds, $L(\mathbf{H_N})$ is compared to $L(\mathbf{H_A})$. If $L(\mathbf{H_A}) > L(\mathbf{H_N})$, the anomaly hypothesis $\mathbf{H_A}$ is accepted, otherwise the null hypothesis if $\mathbf{H_N}$ is accepted. There are two possible outputs: {null, anomaly}. Combining (21) and (22) yields the decision rule for accepting $\mathbf{H_A}$

$$M_+(x, y) > N \cdot \left(\left(p_A^{-1} - 1\right)^{\frac{2}{N+1}} - 1\right), \quad (32)$$

where $M_+(x, y)$ is given by (17), (29), (30), (31).

Monitor 2 (ISO): Fault isolation: Assume that the anomaly hypothesis $\mathbf{H_A}$ holds and $L(\mathbf{H_A}) > L(\mathbf{H_N})$. Then the fault isolation monitor finds a set $\mathcal{J}$ of all input channel fault hypothesis $\mathbf{H_{I,j}}$ such that $L(\mathbf{H_{I,j}}) > L(\mathbf{H_A})$ and a set $\mathcal{K}$ of output channel fault hypothesis $\mathbf{H_{O,k}}$ such that $L(\mathbf{H_{O,k}}) > L(\mathbf{H_A})$. The combined set $\{\mathcal{J}, \mathcal{K}\}$ (the ambiguity group) is the fault isolation monitor output. An empty ambiguity group means the fault is unknown.

The decision rule for accepting input fault hypothesis $\mathbf{H_{I,j}}$ follows from (22) and (23) as

$$M_+(x + f_j z_{I,j}, y) < N \cdot \left((p_F/p_A)^{\frac{2}{N+1}} - 1\right), \quad (33)$$

where $z_{I,j}$ is given by (24). From (22) and (25), the output channel fault hypothesis $\mathbf{H_{O,k}}$ is accepted if

$$M_+(x, y + g_k z_{O,k}) < N \cdot \left((p_F/p_A)^{\frac{2}{N+1}} - 1\right), \quad (34)$$

where $z_{O,k}$ is given by (26). Computing (33) and (34) requires computing (17) and (29), (30), (31).

Monitor 3 (MAP): Most likely hypothesis: Find the hypothesis that has the largest likelihood. The MAP monitor relies on decision rule (32) for accepting anomaly hypothesis $\mathbf{H_A}$. If $\mathbf{H_A}$ is accepted and none of the inequalities (33), (34) hold, then "unknown fault" is reported. If $\mathbf{H_A}$ is accepted and one or more of the inequalities (33), (34) hold then the fault hypothesis corresponding to the smallest left hand side among all inequalities (33), (34) is accepted.

The following parameters are used in the decision rules.

- Matrices Q_N, B_N, S_N that are computed from (29), (30), (31), for training data (5), (6).
- Known input fault signatures f_j in (2) and output fault signatures g_k in (3).
- Prior probabilities p_A, p_F, in (19); these are the tuning parameters and discussed below.
- Regularization parameters p, μ, and ρ in (29), (30), (31) are discussed below in Subsection V-D. In the examples below, parameter p in (29), (30), (31), is $p = m + 1$.

B. Baseline approach as asymptotics

Consider the proposed algorithm in the asymptotic case of a very large training data set, $N \gg 1$.

First, consider the thresholds in the right hand sides of the decision rules (32), (33), (34). The expressions for the thresholds are of the form $\phi(N;\gamma) = N \cdot (\gamma^{\frac{2}{N+1}} - 1)$. It can be proven that for $N, \gamma > 0$ the function $\phi(N;\gamma)$ (i) is monotonically decreasing with N, and (ii) for large N, converges to asymptotic value $\lim_{N\to\infty} \phi(N;\gamma) = 2\log\gamma$. Denote by R the asymptotic value in the right hand side (r.h.s.) of (32); denote by W the asymptotic value in the r.h.s. of (33) and (34), then

$$R = 2\log\left(p_A^{-1} - 1\right), \quad W = 2\log\left(p_F/p_A\right). \tag{35}$$

Second, consider the expression (17) for $M_+(x,y)$, specifically, its denominator. Assume that for $N \to \infty$, we have $Q_N \to Q$ and Q is invertible; then $N^{-1}\|x\|^2_{Q_N^{-1}} \to 0$. This means for $N \gg 1$ we can drop the denominator in (17) and $M_+(x,y) \approx M_1(x,y)$, where $M_1(x,y)$ is of the form (10).

Note that the denominator in (17) might be large even for a large N. This might happen if matrix Q_N is ill-conditioned and x has nonzero projection on its small singular vector (the large singular vector of Q_N^{-1}).

By substituting $M_1(x,y)$ for $M_+(x,y)$ in the left hand sides of (32), (33), (34) and the asymptotic decision thresholds (35) in the right hand sides, we get the asymptotic decision rules to be used for $N \gg 1$. The anomaly hypothesis $\mathbf{H_A}$ is accepted if the following asymptotic version of decision rule (32) holds,

$$\|y - B_N x\|^2_{S_N^{-1}} > R, \tag{36}$$

where R is given by (35).

The input fault isolation condition $L(\mathbf{H_{I,j}}) > L(\mathbf{H_A})$ becomes

$$\|y - B_N(x + z_{I,j} f_j)\|^2_{S_N^{-1}} < W, \tag{37}$$

where $z_{I,j}$ is given by (11) with $h = -B_N f_j$ and W by (35).

Similarly, the output fault isolation condition $L(\mathbf{H_{O,k}}) > L(\mathbf{H_A})$ can be written in the form

$$\|y - B_N x - z_{O,k} g_k\|^2_{S_N^{-1}} < W, \tag{38}$$

where $z_{O,k}$ is given by (11) with $h = g_k$ and W, by (35).

Baseline monitors (10)-(11) use decision rules of the form (36), (37), (38), and are, thus, asymptotically optimal for appropriate decision thresholds.

C. Computational Performance

The monitors of Subsection IV-A are suitable for on-line real time implementation. In on-line monitoring, a new data point $\{x, y\}$ is first processed by a monitor and then added to the historical data set $\mathcal{D}_N$ (1). The computation could be split into two steps. Step 1 is to compute matrices Q_N (29), B_N (30), and S_N (31) from data (5), (6). The computations can be carried recursively, as the new data arrives, by propagating rank-1 updates of the matrices. The computational complexity of Step 1 update is the same as for the well-known Recursive Least Squares update. At Step 2 the hypothesis likelihoods are computed using the results of Step 1 and the new data point $\{x, y\}$. The three monitors of Subsection IV-A rely on decision rules (32), (33), (34). Their computational complexity is quadratic in sizes n of x and m of y.

V. Asymptotic performance and tuning

The asymptotic decision rules in Subsection IV-B depend on two prior probabilities p_A, p_F. In practice, the knowledge of Bayesian priors is often unavailable. The priors could be then considered as tuning parameters. This section described how p_A and p_F can be tuned based on asymptotic performance considerations. A consistent Bayesian formulation uses the same priors for any N.

A. Bayesian parameter tuning

In a Bayesian formulation, the prior probabilities p_A and p_F define the tradeoff between Type I (false positive) and Type II (false negative) errors. They can be considered as tuning parameters. The proposed tuning is based on frequentist performance of the Bayesian decision logic. The use of hybrid Bayesian-frequentist approaches is argued in [9]. Frequentist statistics "... gives a route to assessing methods that may have been suggested on relatively informal grounds" [7].

Tuning the parameters of Bayesian algorithms to achieve desired false alarm performance is discussed in [18], [21]. In [18], the tuning optimizes fault isolation subject to constraint on asymptotic false positive error rates for large faults.

B. Geometric interpretation

Assume that the true covariance S in (4) is a positive definite (invertible) matrix. In accordance with the well known properties of the least square estimator, $S = \lim_{N\to\infty} S_N$ and $B = \lim_{N\to\infty} B_N$.

The covariance S being invertible allows to introduce the whitened residuals

$$r = S^{-1/2}(y - Bx), \tag{39}$$

where $S^{-1/2}$ can be any square root of matrix S. In the coordinates r, decision rules (36), (37), and (38) have simple geometry. The anomaly decision (36) can be written as

$$\|r\|^2 > R. \tag{40}$$

Input fault isolation conditions (37) and output fault isolation conditions (38) have the form

$$\left\|r - (r^T h/\|h\|^2) \cdot h\right\|^2 < W, \tag{41}$$

where $(r^T h)/\|h\|^2 \cdot h$ is the projection of the residual r on the fault signature h. This signature is $h = S^{-1/2} B f_k$ for input fault k and $h = S^{-1/2} g_j$ for output fault j.

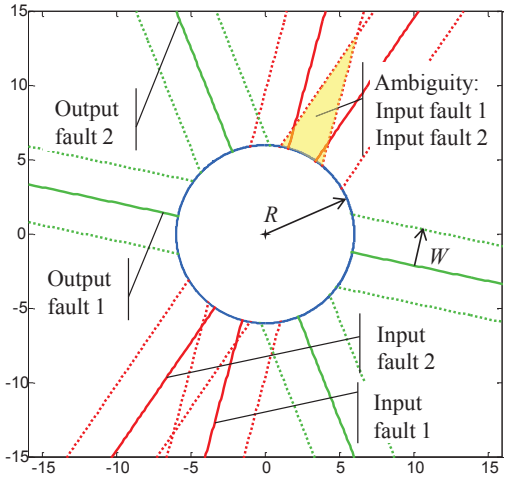

Fig. 1. Geometry of decision rules in a 2×2 example

Figure 1 illustrates decision rules (40) and (41) for an example with $m = 2$. The coordinates are the components of the whitened residual r. The circle (40) of radius R describes the anomaly hypothesis decision. The bands (41) of width W around the fault signature lines describe fault isolation. Monitor 1 (ANO) accepts anomaly hypothesis if r is outside of the circle. Monitor 2 (ISO) includes a fault into the ambiguity group if r is closer than W to the fault signature line. Monitor 3 (MAP) selects the ambiguity group fault with the signature line closest to r.

C. Asymptotic performance

In accordance with (4), the whitened residual r (39) follows the standard normal distribution under the null hypothesis. Therefore, $\|r\|^2 \sim \chi^2_m$.

Decision rules (40) and (41) can be tuned by selecting R and W. Let α be an acceptable asymptotic false positive rate in Monitor 1 (ANO) anomaly detection (36). In accordance with (40), the false positive rate is the probability that $\|r\|^2 > R$. The threshold R should be such that the cumulative distribution function (CDF) for $\|r\|^2$ has value $1 - \alpha$,

$$\mathrm{CDF}_{\chi^2_m}(R) = 1 - \alpha. \tag{42}$$

For Monitor 2 (ISO), W in fault isolation decisions (37) and (38) defines the false negative rate (missing the true fault in the ambiguity group). In accordance with (41), this rate is the probability that $\|r_{\perp h}\|^2 > W$, where $r_{\perp h}$ is the projection of r on the subspace orthogonal to the vector h. Since $\|r\|^2 \sim \chi^2_m$, we have $\|r_{\perp h}\|^2 \sim \chi^2_{m-1}$. The false negative isolation probability α requires that

$$\mathrm{CDF}_{\chi^2_{m-1}}(W) = 1 - \alpha. \tag{43}$$

The described tuning approach has a single tuning parameter α. The thresholds R and W are selected from α in accordance with (42) and (43). This tuning can be generalized to having different target error rates in (42) and (43). A potential downside of such generalization is reduced clarity.

For given R and W, the prior probabilities p_A and p_F (19) can be backed out from (35) as

$$\begin{aligned} p_A &= (1 + e^{R/2})^{-1}, \\ p_F &= e^{W/2}(1 + e^{R/2})^{-1}. \end{aligned} \tag{44}$$

In accordance with (44), we always have $p_F > p_A$. This might seem counter-intuitive since p_A is anomaly prior probability and the anomaly includes all the faults. However, the monitors of Section IV-A evaluate fault hypotheses conditionally on the anomaly hypothesis holding. Therefore, p_F is the fault probability *conditional* on the anomaly occurrence.

D. Finite data performance

Some insights into the algorithm performance for a finite number of the data points N are presented below. In the special case of a single regressor, $n = 1$, $x(t) = 1$, the $m \times 1$ matrix B is the mean vector for y. The anomaly decision rule (33) then corresponds to the well known Hotelling T^2 statistics.

The detection threshold in the right hand side of the anomaly decision rule (33) is larger for smaller N. This reflects the larger uncertainty. In the left hand side of (33), the denominator in $M_+(x, y)$ (17) accounts for the model uncertainty contribution and decreases the anomaly detection sensitivity. The decrease can be substantial if x is in the small singular value subspace of XX^T, i.e., outside of the subspace covered by the training data $\mathcal{D}_N$.

Parameters μ and ρ in (29), (30), (31), do not influence the asymptotic performance. For a finite data set, increasing μ decreases the quadratic form with S_N^{-1} in the numerator of $M_+(x, y)$ (17) thus decreasing the anomaly detection sensitivity. This decrease is suboptimal and small μ is preferable. Increasing ρ reduces the model uncertainty attenuation brought by Q_N^{-1} in the denominator of $M_+(x, y)$ (17). It could increase the error of estimating B_N (7) and, hence, the residual in the numerator of $M_+(x, y)$. A small value of ρ is preferable. The parameters μ and ρ could be chosen based on available prior information or by cross validation.

VI. Verification Study

The developed method was verified through a Monte Carlo study for random systems of a given size.

A. Methodology

In the Monte Carlo study, data set (1) was generated using a 'ground truth' model of the form

$$x(t) = Q_*^{1/2} w(t), \quad (45)$$

$$y(t) = B_* x(t) + S_*^{1/2} v(t), \quad (46)$$

where $v(t)$, $w(t)$ are white noise sequences. Matrices B_*, S_*, and Q_* in (45), (46) were randomly generated from matrix normal distributions $B_* \sim N(0, I_5, I_{10})$, $S_*^{1/2} \sim N(0, I_5, I_5)$, and $Q_*^{1/2} \sim N(0, I_{10}, 4I_{10})$. The fault signatures in (2) and (3) were aligned with the coordinate vectors in the input and output spaces. The signatures were scaled such that $\|B_* f_j\|_{S_*^{-1}} = 1$ and $\|g_j\|_{S_*^{-1}} = 1$.

The study used training data sets (1), (45), (46) with $N = 1000$ points each. For each training set (1), the matrices Q_N, B_N, and S_N in (29), (30), and (31) were computed using the parameters $p = m + 1 = 6$, $\rho = 10^{-4}$, and $\mu = 10^{-4}$. These matrices define the Bayesian decision rules (32), (33), (34) and the Baseline decision rules (36), (37), (38).

For each training data set (1), the decision rules were applied to 16,000 test data samples $\{x, y\}$. These included 1000 samples from the nominal linear Gaussian distribution, 1000 samples for each of the 10 input faults, and 1000 samples for each of the 5 output faults. Faults with six different amplitudes $5.5 \le z_* \le 55.4$ were seeded. Bayesian Monitors 1 (ANO), 2 (ISO), and 3 (MAP) of Subsection IV-A were compared against the Baseline monitors of Subsection IV-B. Figures 2, 3, 4, and 5 show the results averaged over the Monte Carlo runs.

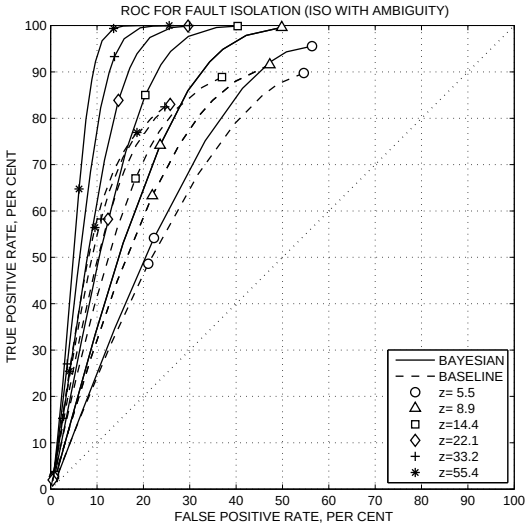

Fig. 2. ROC curve for Monitor 2 (ISO) in the Monte Carlo study

B. Anomaly detection

For fault amplitudes $z_* > 9$ and target error rates $\alpha > 0.001$, the false negative (FN) rate of the anomaly detection was less than 0.01% for both Bayesian and Baseline Monitor 1 (ANO). The observed false positive (FP) rate closely follows the tuning parameter α in (42). For Bayesian Monitor 1 (ANO), FP rate it is just a few percent higher than α because of the finite training set size N. The Baseline ANO monitor has a slightly higher FP rate. For both ANO monitors, the performance is very good. The main practical difference between the Bayesian and Baseline monitors was in the fault isolation performance.

C. Fault Isolation

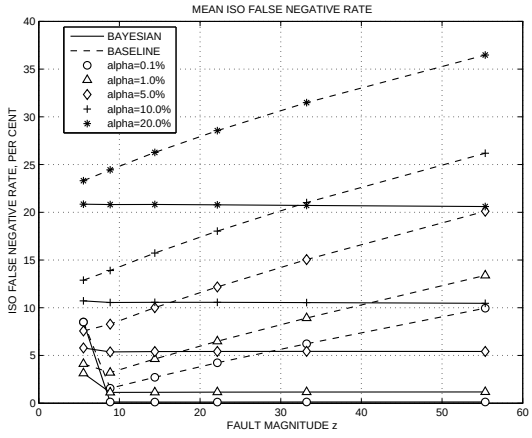

Fig. 3. False negative rates for Monitor 2 (ISO) in the Monte Carlo study

Monitor 2 (ISO) is the main contribution of this paper. Figure 2 shows Receiver Operating Characteristics (ROC) curves for Bayesian and Baseline Monitor 2 (ISO) obtained by varying α in the range $[10^{-12}, 0.95]$. For each of the 15 values of α in the range, the thresholds R and W were set in accordance with (42) and (43) and the priors (19) in accordance with (44).

Each point in Figure 2 plots corresponds to the largest FP and FN rates across all channel faults, averaged over the 1000 Monte Carlo runs. The true positive (TP) rate in the ROC is complementary to the FN rate. FN means that an anomaly was detected, but the ISO monitor did not include the seeded fault into the ambiguity group. FP counts faults that were not seeded yet do appear in the ambiguity group.

Smaller α means larger decision threshold (43) and higher TP and FP rates. Fault magnitudes z_* for the ROC curves are labeled in the plot. As expected, the ROC performance is better for larger z_*. For the same FP rate, the FN rate of the Baseline ISO can exceed the Bayesian ISO FN rate by an order of magnitude or more. For ill-conditioned matrix XX^T, the Bayesian monitor performs well while the Baseline monitor has high error rate.

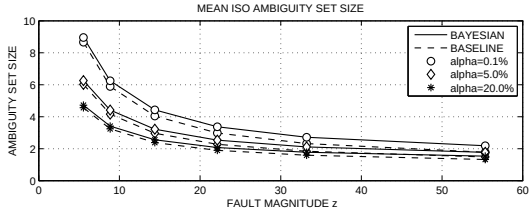

Fig. 4. Ambiguity group size for Monitor 2 (ISO) in the Monte Carlo study

The FN rates for the Baseline and Bayesian ISO are shown in Figure 3. These FN rates were averaged over all channel faults and the Monte Carlo runs. They are plotted vs the fault magnitude z for a few values of the tuning parameter α. With W selected in accordance with (43), the FN rate for the Bayesian ISO monitor closely follows α. The FN rate for the Baseline ISO monitor is substantially higher than that and grows with the fault magnitude. This is because the Baseline ISO monitor is affected by the error in B_N. This error is amplified for large input faults. For $\alpha = 1\%$, the mean FN rate of the Bayesian ISO monitor is an order of magnitude better. For $\alpha = 0.1\%$, it is two orders of magnitude better.

D. Ambiguity group

The average size of the ambiguity group for Bayesian Monitor 2 (ISO) and for Baseline Monitor 2 (ISO) is shown in Figure 4. The averaging is done over all channel faults and the Monte Carlo runs. The plots show the results for a few labelled values of α. As expected, the ambiguity group is larger for smaller z when signal to noise ratio is smaller. The Bayesian ISO has slightly larger ambiguity group compared to the Baseline ISO. This is the price paid for the better performance.

E. MAP Isolation

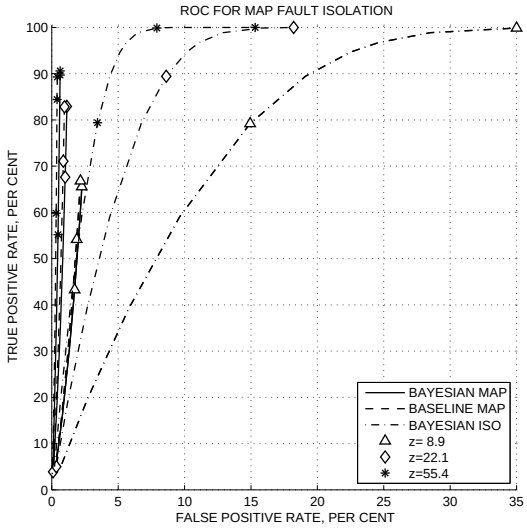

Fig. 5. ROC curves for the Monitor 3 (MAP) in the Monte Carlo study

The ROC curves for Monitor 3 (MAP) fault isolation are shown in Figure 5. These curves show the FP and TP rates averaged across all channel faults and Monte Carlo runs. The ROC curves for Bayesian ISO are included for reference. Where defined, the ROC curves for the MAP monitors are to the left of the Bayesian ISO monitor curves. The results for the Bayesian MAP and the Baseline MAP monitors do not differ much. The mean TP rates for the monitors are below 92% for any α and z_*. For $z_* = 5.5$, the TP rates do not exceed 54%. The TP rates are limited by two factors. (i) As the threshold W increases, the seeded faults are missed less often, but the likelihood of a wrong fault being included increases. (ii) For very small α, the anomaly detection threshold R is so large that the seeded faults do not trigger the anomaly detection. This disables the ISO monitor.

To summarize, the main advantage of the ISO monitor is that it allows to achieve extremely small FN rates (TP rates close to 100%). As discussed above, this is impossible with the MAP monitors. The price for this improvement is higher FP rate related to larger ambiguity group size. The trade-off is described by the ROC curves in Figure 2.

VII. Jet engine monitoring

The proposed approach was applied to the problem of detecting and diagnosing the faults in a jet engine. These safety critical and expensive machines require maintenance action every few thousand hours of usage. Anomaly detection and fault diagnostics are very important for jet engines.

A. Jet engine monitoring problem

Most practical monitoring solutions for the jet engines are based on static models. The industry practice is to collect data samples when engine is in quasi-steady regime during the aircraft takeoff (max power) or cruise. The static performance models are adequate for this. Dynamic models of jet engines are rarely used in practice. The exceptions, such as [41], [42], rely on nonlinear models proprietary to engine manufacturers. A non-proprietary modular nonlinear engine simulation called CMAPSS has been recently developed by NASA, see [27], to support control technology development. It is similar in complexity to the proprietary models. The model in this example is based on CMAPSS-related work [37].

The established approach is to observe the (small) deviations from the nonlinear performance model. The fault isolation algorithms then use the linearization of the nonlinear model for interpreting the deviations. Such linearized model is imperfect. The robust FDI approaches take the modeling error into account, e.g., see [32], but requires to characterize the model uncertainty. This is hard to do in practice.

The proposed approach, which is applied in the example below, is much easier to use. The linearized model is recovered directly from the training data collected in normal operation conditions. The uncertainty of the model estimation is automatically taken into account as described in Section III.

B. Simulated data

Jet engine data are subject to proprietary and other publication restrictions. This paper uses simulated data. To ensure that the results of this paper are reproducible, the simulation model is detailed below. The linearized model of jet engine was based on [37]. This paper assumes that engine deterioration is zero and faults occur at systems inputs and outputs only.

For each data point collected, the engine is assumed to be in the steady state. This assumption holds for engine data sampled in a steady regime. The steady-state deviations of the engine dynamic state from the linearizion point satisfy

$$0 = \bar{A}x + \bar{B}u + w, \tag{47}$$

$$y = \bar{C}x + v, \tag{48}$$

where w is the disturbance and v is the measurement noise. The noises are independent from sample to sample. In (47), (48), the state variables x, the control inputs u, and the observations y describe the deviations from the engine performance model. The components of x and y are explained in Table I. Model parameter matrices $\bar{A}$, $\bar{B}$, $\bar{C}$ have appropriate sizes. The input vectors u, v, and w are independently normally distributed with the standard deviations described by vectors s_u, s_w, and s_v.

Fault	Channel Name	Scale	Units
Inputs (regressor variables) x			
I1	Main burner fuel flow	0.62	lbm/hr
I2	Variable nozzle area	1.76	in^2
I3	Rear bypass door variable area	4.14	in^2
Observations (dependent variables) y			
O1	LPT exit pressure	2.96	psia
O2	LPT exit temperature	13.95	F
O3	Percent LP spool speed	0.42	%
O4	HPC inlet temperature	5.81	F
O5	HPC exit temperature	4.82	F
O6	Bypass duct pressure	0.21	psia
O7	Fan exit pressure	0.09	psia
O8	Booster inlet pressure	0.10	psia
O9	HPC exit pressure	0.93	psia
O10	Core rotor speed	81.01	RPM
O11	LPT blade temperature	16.83	F

TABLE I
MEANING AND SCALES OF THE VARIABLES IN THE ENGINE MODEL

The model parameters taken from [37] are

$$\bar{A} = \begin{bmatrix} 0.9029 & 0.0411 & 0.0381 \\ -0.0069 & 0.9088 & 0.0432 \\ -0.0001 & -0.0004 & 0.9924 \end{bmatrix}, \quad s_w = \begin{bmatrix} 0.3632 \\ 0.6076 \\ 0.0767 \end{bmatrix} \tag{49}$$

$$\bar{B} = \begin{bmatrix} 0.0805 & 0.4928 & -0.1557 \\ 1.0910 & 0.1678 & 0.0341 \\ 0.0018 & -0.0003 & -0.0001 \end{bmatrix}, s_u = \begin{bmatrix} 0 \\ -0.0069 \\ -0.0001 \end{bmatrix} \tag{50}$$

$$\bar{C} = \begin{bmatrix} -0.0034 & 1 & 0.0237 \\ 0.0087 & 0.0002 & 0.0002 \\ 0.0016 & -0.0006 & 0.0001 \\ 0.0022 & -0.0005 & 0.0001 \\ 0.0181 & -0.0024 & 0.0008 \\ 0.0148 & 0.0493 & 0.0094 \\ 0.0018 & 0.0000 & 0.0002 \\ 0.0030 & 0.0127 & 0.0048 \\ -0.0012 & -0.0302 & 0.0656 \\ -0.0172 & -0.1098 & 0.1218 \\ 0.0010 & 0.0007 & 0.0004 \end{bmatrix}, \; s_v = \begin{bmatrix} 0.1933 \\ 13.9400 \\ 0.4231 \\ 5.8080 \\ 4.8255 \\ 0.2066 \\ 0.0889 \\ 0.1010 \\ 0.8506 \\ 81.0133 \\ 16.8429 \end{bmatrix} \tag{51}$$

The simulations in this paper used a closed-loop version of the model (47)–(51). For steady flight, the feedback control adjusts the main fuel flow. The first component u_1 of vector u is such that the fan speed is at the setpoint, $y_3 = 0$. This means u_1 can be computed from $K_3\bar{C} = 0$, where K_3 is the selection matrix for the 3-rd component of y. Then,

$$u_1 = -\frac{K_3\bar{C}(I-\bar{A})^{-1}(\bar{B}_{23}u_{23} + w) + K_1 v}{K_1\bar{C}(I-\bar{A})^{-1}\bar{B}_1}, \tag{52}$$

where $\bar{B}_1$ is the first column of matrix $\bar{B}$, and matrix $\bar{B}_{23}$ is made of the last two columns of $\bar{B}$. The variable nozzle area and the rear bypass door variable area (components of u_{23}) are kept near fixed setpoints. The last two components u_{23} of vector u are independent normal with covariances (50). Fuel flow u_1 controls fan speed y_3 as described by (52).

The described model can be represented by (45), (46).

C. Fault detection and isolation results

Each simulation run produced data $\{x(t), y(t)\}$ in the data set (4). The data driven algorithms of Sections III and IV used matrices X (5) and Y (6) as the training data. Matrices Q_N, B_N, S_N were estimated from the training data. These estimates differ from the ground truth matrices given by (47)–(52). For each training set, the algorithms were tested using multiple data points $x(N+1), y(N+1)$, which included the faults. The simulation, data driven algorithm design, and test data generation were repeated many times to accumulate statistics on the algorithm performance. The algorithm performance evaluation used the ground truth data; the algorithms themselves did not have access to the ground truth.

To verify the obtained decision rules, 1000 test data points $\{x, y\}$ were generated for each training set. Each of 3 faults in input channels and 5 output channel faults were scaled at three magnitudes z_*. The fault signature scales shown in Table I are such that $\|h\|_{S_*^{-1}} = 1$. For each of the $(11+3)\cdot 3 = 42$ fault types, 100 different training data sets were used. For additional 100 training data sets, the test data had no fault. The total was 4,300 training sets with $N = 200$ data points each.

Table II summarizes the results for two monitoring algorithms: Bayesian ISO (Monitor 2) described in Subsection IV-A and Baseline ISO described in Subsection IV-B. Regularization parameters in (29), (30), (31) were $p = m+1 = 12$, $\rho = 10^{-4}$, and $\mu = 10^{-4}$. Bayesian ISO decision rules (32), (33), (34) and Baseline ISO (36), (37), (38) used priors (19) tuned following (44). The decision thresholds $R = 21.3$ and $W = 19.92$ were based on $\alpha = 0.03$ in (42), (43).

Fault	Bayesian ISO False Negative Rate %			Baseline ISO False Negative Rate %		
	$z=5$	$z=8$	$z=15$	$z=5$	$z=8$	$z=15$
FP-A	5.8	5.8	6.0	7.9	7.9	8.1
FN-A	6.0	0.0	0.0	4.4	0.0	0.0
I1	11.7	5.6	5.6	12.0	7.6	8.2
I2	11.3	4.7	4.5	25.4	40.9	89.0
I3	11.6	5.1	4.5	18.8	24.6	66.7
O1	11.5	5.6	5.6	11.8	7.5	7.4
O2	11.6	5.6	5.6	11.9	7.5	7.5
O3	12.0	5.7	5.6	12.2	7.7	7.4
O4	11.6	5.6	5.7	11.9	7.5	7.6
O5	11.4	5.6	5.6	11.7	7.6	7.5
O6	11.4	5.6	5.6	11.7	7.4	7.5
O7	11.5	5.6	5.7	11.7	7.5	7.6
O8	11.6	5.7	5.6	12.0	7.5	7.6
O9	11.4	5.5	5.5	11.8	7.3	7.4
O10	11.4	5.5	5.7	11.7	7.4	7.6
O11	11.4	5.6	5.5	11.7	7.5	7.4

TABLE II
FAULT ISOLATION ERRORS FOR THE ENGINE SIMULATION

The upper section of Table II shows FP-A, false positives for Monitor 1 (ANO) described in Subsections IV-A and IV-B, and FN-A, false negatives for the same anomaly monitors. There are zero false negatives FN-A for $z_* = 8$ and $z_* = 15$ with either Bayesian or Baseline anomaly monitor.

The rest of Table II shows Bayesian ISO error rates for fault magnitudes $z_* = 5$, $z_* = 8$, and $z_* = 15$. The lower two sections of Table II show false negative (FN) rates for the input and output channels described in Table I. The Bayesian algorithm performs well at detecting and identifying all input

and output faults. For faults with $z_* = 8$ and $z_* = 15$, the Bayesian ISO has FN rate around 0.059. This is higher than $\alpha = 0.03$ used in the asymptotic performance tuning in (44), (42), (43), because the training set size is relatively small, $N = 200$. For $N = 10,000$ (the results are not shown), the FN rate of the Bayesian ISO gets close to 0.03. The biggest Bayesian ISO improvement over Baseline ISO is for the input faults I2 "Variable nozzle area" and I3 "Variable bypass door". For larger z_*, an order of magnitude improvement is achieved. The error in the model B_N is amplified for large input faults. The Bayesian algorithm deals with this error. The Baseline algorithm ignores the error, which leads to the error rates in excess of 50% observed for $z = 15$.

For large faults with $z_* = 15$ in inputs I2 and I3, the ambiguity group only occasionally includes more than a single fault. For most other faults with $z_* = 15$, there is no ambiguity for either Bayesian ISO or Baseline ISO algorithms.

VIII. Conclusions

This paper proposed a data-driven method for anomaly detection and fault isolation in linear Gaussian systems. It relies on models estimated from normal training data. The proposed Bayesian formulation takes into account the uncertainty of the estimated models. Currently used anomaly detection and fault isolation methods are limit cases of the proposed formulation for very large training set. The numerical examples show that the proposed method, which includes isolation to an ambiguity group, yields significant improvement of the fault isolation errors. In many cases, the improvement exceeds an order of magnitude. This improvement might come at the price of spurious faults being included in the ambiguity group. The proposed monitoring method can be implemented recursively with the computational cost comparable with the recursive least squares update.

References

[1] Alcala, C.F. and Qin, S.J., "Reconstruction-based contribution for process monitoring," *Automatica*, Vol. 45, No. 7, 2009, pp. 1593–1600.

[2] Basseville, M., "An invariance property of some subspace-based detection algorithms," *IEEE Trans. on Signal Processing*, Vol. 47, No. 12, 1999, pp. 3398–3400.

[3] Bassily, H., Lund, R., and Wagner, J. "Fault detection in multivariate signals with applications to gas turbines," *IEEE Trans. on Signal Processing*, Vol. 57, No. 3, pp. 835–842.

[4] Chen, J. and Patton, R.J., *Robust Model–Based Fault Diagnosis for Dynamic Systems,* Kluwer Academic Publishers, 1999.

[5] Chu, E., Gorinevsky, D., and Boyd, S., "Scalable statistical monitoring of fleet data," *18th World IFAC Congress*, August 2011, Milano, Italy.

[6] Couillet, R. and Debbah, M., "Signal processing in large systems: A new paradigm," *IEEE Signal Process. Mag.*, Vol. 30, No. 1, 2012, pp. 24–39.

[7] Cox, D.R., "Frequentist and Bayesian statistics: A critique," *PHYSTATO5: Statistical Problems in Particle Physics, Astrophysics and Cosmology*, pp. 3-6, Ed. L. Lyons and M.K. Unel., London, England, Imperial Coll. Press, 2006.

[8] Chu, E., Gorinevsky, D. and Boyd, S, "Detecting aircraft performance anomalies from cruise flight data," *AIAA Infotech@Aerospace Conf.*, April 2010, Atlanta, GA, AIAA-2010-3308.

[9] Efron, B., "Bayesians, Frequentists, and Scientists," *J. American Statistical Assoc.*, Vol. 100, No. 469, 2005, pp. 1-5.

[10] Frank, P.M., Ding, X., "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *Journal of Process Control*, Vol. 7, No. 6, 1997, pp.403–424.

[11] Ge, Z., L. Xie, U. Kruger and Z. Song (2012). "Local ICA for multivariate statistical fault diagnosis in systems with unknown signal and error distributions," *AIChE Journal*, Vol. 58, No. 8, pp. 2357–2372.

[12] Gorinevsky, D., Matthews, B., and Martin, R., "Aircraft anomaly detection using performance models trained on fleet data," *Conf. on Intelligent Data Understanding (CIDU)*, October 2012, Boulder, CO.

[13] Gorinevsky, D., "Bayesian fault isolation in multivariate statistical process monitoring," *American Control Conf.*, June 2011, San Francisco, CA.

[14] Gorinevsky, D., Boyd, S., and Poll, S., "Estimation of faults in DC electric power system," *American Control Conf.*, June 2009, St. Louis, MO

[15] Hotelling, H., "The generalization of Student's ratio," *Annals of Mathematical Statistics*, Vol. 2, No. 3, 1931, pp.360–378.

[16] Draper, D., "Assessment and propagation of model uncertainty (with discussion)," *Journ. of the Royal Statistical Society. Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. 45–97.

[17] Dunia, R. and Qin, S.J., "Subspace approach to multidimensional identification and reconstruction of faults," *AIChE Journ.*, Vol. 44, No. 8, 1998, pp. 1813–1831.

[18] Fillatre, L. and Nikiforov, I. "Asymptotically uniformly minimax detection and isolation in network monitoring," *IEEE Trans. on Signal Processing*, Vol. 60, No. 7, 2012, pp. 3357-3371

[19] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. *Bayesian Data Analysis*, 2nd Ed., Chapman and Hall, 2003.

[20] Gupta, A.K. and Najar, D.K., *Matrix Variate Distributions*, Chapman and Hill, 1999.

[21] Baygun, B. and Hero, A.O., ' "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. on Information Theory*, Vol. 41, No. 3, 1995, pp. 688–703.

[22] Hwang, I., Kim, S., Kim, Y., and Seah, C.E., "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, Vol. 18, No. 3, 2010, pp. 636–653.

[23] Kelly, E.J., "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 22, No. 2, 1986, pp. 115–127.

[24] Kourti, T. and MacGregor, J.F., "Multivariate SPC methods for process and product monitoring," *Journ. of Quality Technology*, Vol. 28, No. 4, 1996, pp. 409-427.

[25] Li, G., Alcala, C. F., Qin, S. J., and Zhou, D., "Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the Tennessee Eastman process," *IEEE Trans. on Control Systems Technology*, vol.PP, no.99, pp.14, doi: 10.1109/TCST.2010.2071415

[26] MacGregor, J.F., Yu, H., Munoz, S.G., and Flores-Cerrillo, J.,"Data-based latent variable methods for process analysis, monitoring and control ," In *Computers and Chemical Engineering*, Vol 29, No. 6, 2005, pp. 1217–1223.

[27] May, R.D., Csank, J., Lavelle, T.M., Litt, J.S., and Guo, T.H., "A high-fidelity simulation of a generic commercial aircraft engine and controller," *46th AIAA Joint Propulsion Conference*, Nashville, TN, 2010. AIAA20106630,

[28] Middleton. D. and Esposito, R., "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. on Information Theory*, Vol. 14, No. 3, 1968, pp. 434–444.

[29] Montgomery D.C. *Introduction to Statistical Quality Control*, (4th edn). Wiley: New York, 2000.

[30] Nevat, I., Peterst, G.W., and Yuan, J., "Bayesian inference in linear models with a random gaussian matrix: algorithms and complexity," *IEEE 19th Internat. Symp. on Personal, Indoor and Mobile Radio Comm.*, Sept. 2008, Cannes, France.

[31] Nomikos, P and MacGregor J. F., "Multi-way partial least squares in monitoring batch processes," *Chemometrics and Intelligent Laboratory Systems*, Vol. 30, 1995, pp. 97-108.

[32] Patton, R.J. and Chen, J. "Uncertainty modeling and robust fault diagnosis for dynamic systems," In *Issues of Fault Diagnosis for Dynamic Systems*, ed. R.J. Patton, R.N. Clark, and P.M. Frank, pp. 189–218, Springer, London, 2000.

[33] Piovoso, M.J., Kosanovich, K.A., and Yuk, J.P., "Process data chemometrica," *IEEE Trans. on Instrumentation and Measurement*, Vol. 41, No. 2, 1992, pp. 262–268.

[34] Raich, A. and Cinar, A., "Diagnosis of process disturbances by statistical distance and angle measures," *Computers and Chemical Engineering*, Vol. 21, No. 6, 1997, pp. 661–673.

[35] Runger, G.C., Alt, F.B., and Montgomery, D.C., "Contributors to a multivariate statistical process control chart signal", *Communications in Statistics - Theory and Methods,* Vol. 25, No. 10, 1996, pp. 2203–2213.

[36] Simani, S., Fantuzzi, C., and Beghelli, S., "Diagnosis Techniques for Sensor Faults of Industrial Processes," *IEEE Trans. on Control Systems Technology*, Vol. 8, No. 5, 2000, pp. 848–855.
[37] Simon, D. and Simon, D.L., "Constrained Kalman filtering via density function truncation for turbofan engine health estimation," *Int. J. of Systems Science*, Vol. 41, No. 2, 2010, pp. 159-171.
[38] Kwan, C., and Xu, R., "A note on simultaneous isolation of sensor and actuator faults," *IEEE Trans. on Control Systems Technology*, Vol. 12, No. 1, 2004, pp. 183–192.
[39] Simon, D.L. and Armstrong, J.B. "An integrated approach for aircraft engine performance estimation and fault diagnostics," *J. Eng. Gas Turbines Power*, Vol. 135, No. 7, 071203, 2013. Paper No: GTP-12-1447
[40] Van Huffel, S. and Lemmerling, P., *Total Least Squares and Errors-In-Variables Modeling: Analysis, Algorithms and Applications*, Springer-Verlag, 2002. London, UK.
[41] Viassolo, D.E., Adibhatla, S., Brunell, B.J., Down, J.H., Gibson, N.S., Kumar, A., Mathews, H.K., and Holcomb, L.D., "Advanced estimation for aircraft engines," *American Control Conference*, June 2007, Pages 2807 - 2821.
[42] Volponi, A.J., DePold, H., Ganguli, R., and Daguang, C., "The use of Kalman Filter and neural network methodologies in gas turbine performance diagnostics: A comparative study," *J. Eng. Gas Turbines Power*, Vol. 125, No. 4, 2003, pp. 917-924.
[43] Wiesel, A., Eldar, Y.C., and Yeredor, A., "Linear regression with Gaussian model uncertainty: algorithms and bounds," *IEEE Trans. on Signal Processing*, Vol. 56, No. 6, 2008, pp. 2194-2205
[44] Zeng, J., L. Xie, U. Kruger and C. Gao (2014). "Regression-based analysis of multivariate non-Gaussian datasets for diagnosing abnormal situations in chemical processes," *AIChE Journal*, Vol. 60, No. 1, pp. 148-159.
[45] Zhang, X., Tang, L., and Decastro, J., "Robust fault diagnosis of aircraft engines: A nonlinear adaptive estimation-based approach," *IEEE Trans. on Control Systems Technology*, Vol. 21, No. 3, 2013, pp. 861–868.
[46] A. Zymnis, S. Boyd, and D. Gorinevsky, "Relaxed maximum a posteriori fault identification," *Signal Processing*, Vol. 89, No. 6, 2009, pp. 989–999.

APPENDIX A: POSTERIORS FOR LINEAR REGRESSION

Using the notation (5), (6), MLE problem (12), (13) can be written in the form

$$L_N = \max_{B,S} \left[-\frac{1}{2}\text{trace } \left((Y-BX)^T S^{-1}(Y-BX)\right) - \frac{N}{2}\log\det(S) \right]. \quad (53)$$

One can verify that $B = B_N$ (7) and $S = S_N$ (8) solve the normal equations for (53)

$$S^{-1}[YX^T - BXX^T] = 0, \quad (54)$$

$$\frac{1}{2}NS^{-1} - \frac{1}{2}S^{-1}(Y-BX)(Y-BX)^T S^{-1} = 0. \quad (55)$$

From (55) and trace properties one can see that

$$L_N(B_N, S_N) = -\frac{N}{2}\log\det(S_N) - \frac{Nm}{2}. \quad (56)$$

Substituting S_N (8) and $B = B_N$ (7) into (56) yields (14).

The solution of the extended problem (15) can be expressed through (13) by substituting

$$X \leftarrow X_+ = [X \;\; x], \quad X \leftarrow Y_+ = [Y \;\; y], \quad N \leftarrow N+1.$$

From (8) and (7) the covariance S_+ in the extended problem satisfies

$$(N+1)S_+ = Y_+Y_+^T - Y_+X_+^T(X_+X_+^T)^{-1}X_+Y_+^T, \quad (57)$$

where $Y_+Y_+^T = YY^T + yy^T$, $Y_+X_+^T = YX^T + yx^T$, and $X_+X_+^T = XX^T + xx^T$. From Inverse Matrix Lemma

$$(X_+X_+^T)^{-1} = (XX^T)^{-1} - \frac{(XX^T)^{-1}xx^T(XX^T)^{-1}}{1+x^T(XX^T)^{-1}x}.$$

Using the notation $r_x = x^T(XX^T)^{-1}x$, $u = YX^T(XX^T)^{-1}x = B_Nx$ and grouping terms we get

$$(N+1)S_+ = YY^T + yy^T - YX^T(XX^T)^{-1}XY^T - \left(-\frac{uu^T}{1+r_x} + \frac{yu^T}{1+r_x} + \frac{uy^T}{1+r_x} + \frac{yy^Tr_x}{1+r_x}\right). \quad (58)$$

This can be expressed as

$$S_+ = \frac{N}{N+1}S_N + \frac{1}{N+1}\cdot\frac{(u-y)(u-y)^T}{1+r_x}. \quad (59)$$

Applying Matrix Determinant Lemma to (59) yields

$$\det(S_+) = \left[1 + \frac{\|u-y\|^2_{S_N^{-1}}}{N(1+r_x)}\right]\det(S_N)\left[\frac{N}{N+1}\right]^m, \quad (60)$$

where the notation $\|q\|^2_{S_N^{-1}} = q^TS_N^{-1}q$ is used.

The version of (56) for the extended problem (15) is

$$L_+ = \frac{N+1}{2}\log\det(S_+) + \frac{(N+1)m}{2}.$$

Substituting (60) into the formula for L_+ results in the log posterior distribution for $y|x, Y, X$, a direct precursor of (16), (17)

$$L_+ = -\frac{N+1}{2}\log\left(1 + \frac{\|y-B_Nx\|^2_{S_N^{-1}}}{N(1+x^T(XX^T)^{-1}x)}\right) - \frac{N+1}{2}\log\det(S_N) - \frac{(N+1)m}{2}\log\frac{N}{N+1} - \frac{(N+1)m}{2}.$$

Dimitry Gorinevsky (M'91–SM'98–F'06) received a Ph.D. from Department of Mechanics and Mathematics at Moscow (Lomonosov) University in 1986, and a M.Sc. from Moscow Institute of Physics and Technology (Phystech) in 1982. He is a Consulting Professor in Electrical Engineering at Stanford University and a principal of Mitek Analytics LLC consultancy in Palo Alto, CA. In the past, he spent 10 years with Honeywell. He worked on decision and control systems application across several industries. His current interests are in Big Data analytics. Dr. Gorinevsky was an Associate Editor of the IEEE Transactions on Control Systems Technology, 2001-2008. He received Control Systems Technology Award, 2002, and Transactions on Control Systems Technology Outstanding Paper Award, 2004, of the IEEE Control Systems Society. He received Best Paper Award (Senior Award), 2013, of the IEEE Signal Processing Society. He is a Fellow of IEEE.