Fault detection in non-linear systems based on GP-EKF and GP-UKF algorithms

Behrooz Safarinejadian & Elham Kowsari


To link to this article: https://doi.org/10.1080/21642583.2014.956843
Fault detection in non-linear systems based on GP-EKF and GP-UKF algorithms

Behrooz Safarinejadian* and Elham Kowsari

Control Engineering Department Shiraz University of Technology Modarres Blvd., P.O. Box 71555-313 Shiraz, Iran
(Received 22 April 2014; accepted 18 August 2014)

In this paper, two new fault detection methods are proposed for non-linear systems. The proposed methods are based on combining an extended Kalman filter (EKF) and an unscented Kalman filter (UKF) with Gaussian processes (GPs). One of the major advantages of these algorithms is that they do not need the system's model while they have an accurate and fast operation in fault detection. In order to show the promising performance of the proposed algorithms, they are applied to an aeroplane tracking system with a highly non-linear dynamics. Superiority of the GP-UKF over GP-EKF in fault detection is also shown based on the simulation results.

Keywords: fault detection; unscented Kalman filter; extended Kalman filter; Gaussian processes; state estimation

1. Introduction

Advances in science and technology have resulted in larger and more complex systems in which fault occurrence is inevitable. Since proper and faultless operation of these systems is the main concern of industrial system designers, a fault detection and diagnosis unit is an important part of these systems. Nowadays, detection of faults in the shortest time and with the least cost is vital in industrial systems. After detecting a fault, a proper control operation can be done for fault diagnosis.

The methods of fault detection in dynamic systems can be divided into two general categories. First, methods in which a system model is not required and the process of fault detection is carried out just using the measured data. These methods need a large data set for proper operation, which is their main weak point. Statistical and expert methods belong to this category (Isermann, 2006; Safarinejad, Ghane, & Monirvaghefi, 2013). Second, methods in which a mathematical model of the system is required for fault detection (Isermann, 2006). In these methods, the output of the system is compared with the output of a reference faultless system and the resulting residual signal is used for fault detection. Since estimation algorithms are usually used in these methods to estimate the states of the system, having a proper model of the system and its states is required for fault detection.

Various methods of state estimation have been proposed for dynamic systems up to now. In 1960, Kalman filter (KF) was introduced as an optimal solution for state estimation in linear systems (Kalman, 1960; Safarinejad & Mozafrari, 2013). Due to the high computational demand of KF in some situations, information filter (IF) was proposed afterwards (Anderson & Moore, 1979). The main difference between KF and IF is that KF propagates the state vector and its corresponding covariance matrix using the system dynamics, whereas IF propagates the information vector which is the inverse of the covariance matrix. The computational demand of the IF is less than that of the KF when the number of outputs is larger than the order of the system (Simon, 2006). However, these algorithms are limited to linear systems, although most of the practical systems are non-linear in nature. Since KF cannot estimate the states of a non-linear system, suboptimal estimation approaches should be used instead. Extended Kalman filter (EKF) is a well-known suboptimal method in which process and measurement equations are linearized (Ljung, 1979). In this filter, probability density function (pdf) of the states is assumed to be Gaussian. Linearization errors of EKF may cause the estimation accuracy to become low or even may result in filter divergence. Furthermore, due to the Jacobian matrix computation that is required for the linearization process, computational load of EKF is relatively high. Extended IF (EIF) was proposed to solve this problem of EKF (Mutambara, 1998). However, large linearization errors of EKF and EIF, especially in highly non-linear systems, have caused deficiency of these filters. To solve these problems, another suboptimal approach named unscented Kalman filter (UKF) was proposed in which a collection of sample points (sigma points) is used for mean and covariance propagation (Julier, Uhlmann, & Durrant-Whyte, 2000). Using these sigma points decreases the computational volume (Julier & Uhlmann, 2004).
Model-based methods are a class of well-known fault detection approaches in which more accurate estimates result in a better fault detection performance. Therefore, researchers have used parametric methods, which reveal a relatively accurate estimate for fault detection (Del Gobbo, Napolitano, Famouri, & Innocenti, 2001; Efimov, Zolghadri, & Simon, 2011, 2013; Foo, Zhang, & Vilathgamuwa, 2013; Safarinejadian, Menhaj, & Karrari, 2009; Zhang et al., 2013). EKF was used for fault detection and also robust fault detection (Zarei and Poshtan, 2011). UKF was also used for fault detection in non-linear systems (Alkov and Weidemann, 2013; Liu, Liu, Lu, & Wang, 2014; Shang and Liu, 2011; Xiong and Zhang, 2005; Yuan and Ye, 2013).

The aforementioned filters use a parametric model of the system for state estimation. Even though the performance of these filters is acceptable, reaching an accurate parametric model of the system is difficult. Furthermore, since considering all aspects of modelling is difficult in practice, a simple parsimonious model is usually used. Therefore, the prediction capability of these models will also be limited. In order to overcome the aforementioned limitations of the parametric state estimation methods, Gaussian process (GP) regression models were proposed for process and measurement equations learning (Rasmussen, 2006). GP was used for dynamic system identification and prediction in Rasmussen (2006). Recently, by combining parametric and non-parametric methods, new algorithms are proposed for state estimation in non-linear systems (Deisenroth, Huber, & Hanebeck, 2009; Ko & Fox, 2011; Reece & Roberts, 2010).

GPs have also been used for fault detection in dynamic systems (Juricic & Kocijan, 2006; Juricic, Ettler, & Kocijan, 2011; Serradilla, Shi, & Morris, 2011). Due to the aforementioned problems in the modelling of non-linear systems and the need for accurate state estimates for fault detection in non-linear systems, a residue is required that is very sensitive to fault presence and instantaneously detects the fault occurrence.

In this paper, two new methods of fault detection will be proposed for non-linear systems. The proposed methods are based on combining EKF and UKF with GPs. The main advantage of combining the GPs and the existing filters such as EKF and UKF is that EKF and UKF need the exact system’s model, while the proposed approaches do not need any model of the system. In other words, the existing filters such as EKF and UKF are model-based methods, whereas the two proposed methods (GP-EKF and GP-UKF) are non-model-based approaches that do not require an accurate model of the system while their generated residue results in an accurate fault detection performance.

It should be noticed that the algorithms proposed in this paper provide a new non-parametric approach for fault detection. Artificial neural network (ANN) can be considered as the main class of traditional non-parametric fault detection methods in which an input–output data set is the only information required about the dynamic system (Isermann, 2006).

ANN is a black-box system identification method in which the unknown system is approximated by an ANN. In this method, the output of the trained ANN is only a deterministic value corresponding to its input data (Isermann, 2006). However, using the proposed GP-EKF and GP-UKF, the model approximated for the unknown function provides a Gaussian pdf as an output in which the mean and covariance are determined by the GP regression method.

Furthermore, using ANN for fault detection is rather difficult in practice. Selecting the structure of the ANN, the number of its neurons, the activation function, the learning rate, etc., make using ANN difficult in practice. Moreover, there is not any systematic method for selecting these items. However, using the GP regression method, a specified Gaussian distribution is chosen for the unknown function in which the mean and covariance are determined by the trained data set (Rasmussen, 2006).

The objectives of this paper include (1) proposing two novel non-parametric methods in fault detection that provide accurate residue signal and (2) applying these methods for fault detection in an aircraft tracking system as a strictly non-linear system.

The rest of the paper is organized as follows. In Section 2, problem formulation is proposed. EKF, UKF and GPs are introduced briefly in this section. In Section 3, the proposed algorithms are given and how they can be used in fault detection is discussed. Section 4 is devoted to the simulation results. Finally, Section 5 concludes the paper.

2. Problem formulation

In this section, GP regression, EKF and UKF are introduced briefly.

2.1. Gaussian regression model

GPs are powerful non-parametric tools for learning unknown functions using a training data set. This data set includes input–output pairs and the GPs give a mapping between inputs and outputs. The main characteristic of GPs is their flexibility in modelling, since it can model the system’s behaviour in the presence of uncertainties. Furthermore, GPs can estimate the input noise and smooth the parameters using the training data set (Boyle, 2007; Rasmussen, 2006; Shi & Choi, 2011).

Definition A GP is a collection of random variables, any GP finite number which has a joint Gaussian distribution (Rasmussen, 2006).

In the following, GP will be expressed in brief. Consider a training data set \( D_f = \{ X, y \} \) in which \( X = \{ x_1, x_2, \ldots, x_d \} \) contains \( d \)-dimensional inputs \( x \), and \( y = \{ y_1, y_2, \ldots, y_n \} \) contains observations or targets \( y \). It is
assumed that the training data set \( D_f \) is obtained from
\[
y_i = f(x_i) + \varepsilon, \tag{1}
\]
where \( \varepsilon \) is a zero mean Gaussian noise with variance \( \sigma_n^2 \), that is \( \varepsilon \sim N(0, \sigma_n^2) \).

GP regression tries to approximate the function \( f \) using the training data set \( D_f \). Just like a random variable with a Gaussian distribution that can be characterized solely by its mean and covariance, it is sufficient to find a proper mean and covariance function for modelling an unknown function with GP.

The mean function that is used for GP based on a desired input \( x_* \) is given as
\[
\text{GP}_y(x_*, D_f) = k_*^T K^{-1} y. \tag{2}
\]

The variance function of GP with the same input can also be defined as
\[
\text{GP}_\Sigma(x_*, D_f) = k(x_*, x_*) - k_*^T K^{-1} k_*, \tag{3}
\]
where \( K \) is an \( n \times n \) matrix of kernel functions between the training inputs. A kernel function type is a designer’s choice but it usually has a square exponential form added with a noise term as
\[
k(x, x') = \sigma_f^2 e^{-1/2(x-x')^TW(x-x')^T} + \sigma_n^2 \delta, \tag{4}
\]
where \( \sigma_f^2 \) is the variance of the signal that controls the prediction uncertainty in the regions in which the density of the training data is low. \( W \) is a diagonal matrix which contains length scales of process such that \( W = \text{diag}[1/l_1^2, 1/l_2^2, \ldots, 1/l_n^2] \). \( \delta \) denotes the Kronecker delta function and \( \sigma_n^2 \) controls the process noise. In Equation (3), \( K[i, j] = k(x_i, x_j) \). \( k_* \) is a vector defined by the kernel values between \( x_* \) and the training inputs \( X \), given in the following equation:
\[
k_*[i] = k(x_*, x_i). \tag{5}
\]

### Learning of the hyperparameters

In order to obtain a proper GP model that optimally approximates the unknown function \( f \), the parameters of the GP, that is, the parameters of the kernel function, should be obtained optimally. For this purpose, a hyperparameter vector \( \theta = [W, \sigma_f, \sigma_n] \) is considered that is determined by maximizing the log marginal likelihood of training outputs given inputs as follows (Quinonero-Candela, Rasmussen, & Williams, 2007):
\[
\theta_{\text{max}} = \arg \max \{ \log(p(y | X, \theta)) \}. \tag{6}
\]

The logarithmic term of Equation (6) can be written as
\[
\log(p(y | X, \theta)) = -\frac{1}{2}y^T (K(X, X) + \sigma_n^2 I)y - \frac{1}{2} \log |K(X, X) + \sigma_n^2 I| - \frac{n}{2} \log 2\pi. \tag{7}
\]

This optimization problem can be solved using different methods such as conjugate gradient. The partial derivatives that are required for optimization are given as
\[
\frac{\partial}{\partial \theta_k} \log(p(y | X, \theta)) = \frac{1}{2} \text{tr} \left[ (K^{-1})_y (K^{-1})^T \frac{\partial K}{\partial \theta_k} \right], \tag{8}
\]
\[
\frac{\partial k(i, j)}{\partial \sigma_f} = 2\sigma_f e^{-1/2(x_i - x_j)^TW(x_i - x_j)^T},
\]
\[
\frac{\partial k(i, j)}{\partial \sigma_n} = 2\sigma_n \delta,
\]
\[
\frac{\partial k(i, j)}{\partial W_{ii}} = -\frac{1}{2} (x_i[i] - x_i[j])^2 \times \sigma_f^2 e^{-1/2(x_i - x_j)^TW(x_i - x_j)^T}. \tag{9}
\]

### 2.2. Extended Kalman filter and unscented Kalman filter

In the procedure of non-linear system state estimation, there are non-linear integrals that have no closed-form solution. Therefore, researchers have tried to use sub-optimal methods for this problem (Anderson & Moore, 1979; Ljung, 1979). Linearizing the non-linear systems and numerical integration are two suboptimal methods that have been considered for this problem and have resulted in the proposal of EKF and UKF, respectively. These two filters will be discussed in Sections 2.2.1 and 2.2.2 in brief.

#### 2.2.1. Extended Kalman filter

In the case that the state and measurement equations are non-linear, an approximate solution is to linearize these equations using the Taylor series expansion around the mean of the Gaussian random variable (GRV). Afterwards, the standard KF is applied to this linearized model. The resulting filter is called extended KF (EKF). EKF is not an optimal filter and its performance depends on the linearization accuracy. Process and measurement noises are assumed to be Gaussian and the accuracy of this filter is of the first order. In the EKF, process and measurement equations are in the general form of
\[
x_k = f_{k-1}(x_{k-1}) + v_{k-1}, \tag{10}
\]
\[
z_k = h_k(x_k) + w_k, \tag{11}
\]
where \( f \) and \( h \) are non-linear functions, \( x_k \) is the state vector and \( v_{k-1} \) and \( w_k \) are process and measurement noises,
respective. Partial differentiation is used for linearization as follows:

\[ F[k - 1] = \frac{\partial f}{\partial x_{k-1}} \bigg|_{x_{k-1} = \hat{x}_{k-1|k-1}}, \tag{12} \]

\[ H[k] = \frac{\partial h}{\partial x_{k}} \bigg|_{x_{k} = \hat{x}_{k|k-1}} \tag{13} \]

The EKF equations are given as

\[ x_{k|k-1} = f_{k-1}(\hat{x}_{k-1|k-1}), \tag{14} \]

\[ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_{k}(x_{k} - h_{k}(\hat{x}_{k|k-1})), \tag{15} \]

\[ P_{k|k-1} = Q_{k-1} + F[k - 1]P_{k-1|k-1}F[k - 1]^T, \tag{16} \]

\[ P_{k|k} = (I - K_{k}H[k])P_{k|k-1}, \tag{17} \]

where the Kalman gain is defined as

\[ K_{k} = P_{k|k-1}H[k]^T(R_{k} + H[k]P_{k|k-1}H[k]^T)^{-1}. \tag{18} \]

### 2.2.2. Unscented Kalman filter

One of the suboptimal methods in state estimation is based on numerical integration. In this method, a minimum set of weighted sigma points is selected and then the integral is approximated using these sigma points. The sigma points are generated based on the a priori mean and covariance of the random variables. As these points pass through the non-linear system function, the a posteriori mean and covariance of the random variables are obtained. In this filter, the position of the sigma points and their weights are calculated based on the fact that the selected system sigma points should represent the main statistical characteristics of an a priori random variable (Julier & Uhlmann, 2004). In order to reach this goal, UKF utilizes the unscented transform (Julier et al., 2000). The number of sigma points is \( r = 2n_{s} + 1 \), where \( n_{s} \) represents the number of state variables (or the dimension of the state vector) (Van Der Merwe, 2004; Julier & Uhlmann, 2004). The UKF algorithm is given briefly in Figure 2.

Considering the non-linear system represented by Equations (10) and (11), the sigma points and their corresponding weights are given by

\[ \chi_{0} = \tilde{x}, \tag{19} \]

\[ \chi_{i} = \tilde{x} + (\sqrt{n_{s} + \lambda}P_{i}), \quad i = 1, \ldots, n_{s}, \tag{20} \]

\[ \chi_{i} = \tilde{x} - (\sqrt{n_{s} + \lambda}P_{i}), \quad i = n_{s} + 1, \ldots, 2n_{s}, \tag{21} \]

\[ W_{0}^{m} = \frac{\lambda}{(n_{s} + \lambda)}, \tag{22} \]

\[ W_{0}^{m} = \frac{\lambda}{(n_{s} + \lambda)} + (1 - \alpha^{2} + \beta), \tag{23} \]

\[ W_{0}^{m} = W_{i} = \frac{1}{2(n_{s} + \lambda)}, \quad i = 1, \ldots, 2n_{s}, \tag{24} \]

where \( \lambda = \alpha^{2}(l + k) - n_{s} \) is the scaling parameter. \( \alpha \) has a small positive value (usually \( 10^{-3} \leq \alpha \leq 1 \)) and represents the spread of sigma points around \( \tilde{x} \). \( k \) is the second calibration parameter and usually has a value of 0 or 3. \( n_{s} \) is a scalar parameter used to incorporate any extra prior knowledge of the distribution of \( x \) (for a normal distribution, \( \beta = 2 \) is optimal, Wan, Van Der Merwe, & Nelson, 1999). \( \sqrt{(n_{s} + \lambda)P_{i}} \) is the \( i \)th weighted column of square root of the covariance matrix \( P_{i} \).

### 3. The proposed algorithms for fault detection

As discussed earlier, the main disadvantage of the Bayesian filters is that they require the state and measurement equations to be known. To solve this problem, two algorithms will be proposed by combining EKF and UKF with GPs. Afterwards, these algorithms will be used for fault detection.

#### 3.1. GP incorporated with EKF

A new GP-EKF algorithm is proposed here for fault detection in non-linear systems. In this method, process and measurement equations \( f \) and \( h \) are approximated separately using GPs and then the GP models obtained for these functions are incorporated with the EKF algorithm. Eventually, a proper estimate of the system states will be revealed. Covariance matrices of the process and measurement noises, \( Q \) and \( R \), are obtained using the training data set. The process model maps the state and input variables \( (x_{k-1}, u_{k}) \) into a state transition \( \Delta x_{k} = x_{k+1} - x_{k} \). The measurement model maps the state \( x_{k} \) into a measurement \( z_{k} \). In order to obtain these two models, we need two separate training data sets such as

\[ D_{f} = \{(X, U), X'\}, \]

\[ D_{h} = \{X, Z\}, \tag{25} \]

where \( X \) is a matrix of state and \( X' = [\Delta x_{1}, \Delta x_{2}, \ldots, \Delta x_{n}] \) is the matrix containing states transition. \( Z \) is a matrix of the observed inputs. The GP approximation of the \( f \) and \( h \) functions can be denoted by GP\( f \) and GP\( h \), respectively.

In Bayesian filters, the aim of the prediction step is to determine the pdf of \( x_{k} \), that is, \( p(x_{k} | z_{1:k-1}) \), which is done by means of the following equation:

\[ p(x_{k} | z_{1:k-1}) = \int p(x_{k} | x_{k-1}, u_{k-1}) p(x_{k-1} | z_{1:k-1}) dx_{k-1}, \tag{26} \]

where \( p(x_{k-1} | z_{1:k-1}) \) is a normal pdf computed at the previous iteration of the filter. \( p(x_{k} | x_{k-1}, u_{k-1}) \) can be obtained using Equation (10). It has been assumed in this paper that \( f_{k-1} \) can be approximated by a GP in which the mean and covariance are represented by Equations (2) and
(3), respectively. Therefore,

\[ p(x_k | x_{k-1}, u_{k-1}) \approx \mathcal{N}(GP_{\mu}(x_{k-1}, u_{k-1}), D_f), \]

\[ GP_{\Sigma}(x_{k-1}, u_{k-1}, D_f)). \]  

(27)

Thus, \( p(x_k | z_{1:k-1}) \), computed by Equation (26), can be approximated by a normal pdf. In consequence\( x_k = GP_{\mu}(x_{k-1}, u_{k-1}), D_f) + \epsilon_k, \)

\[ \epsilon_k \sim \mathcal{N}(0, GP_{\Sigma}(x_{k-1}, u_{k-1}, D_f)). \]  

(28)

In the update step of the Bayesian filter, the aim is to determine \( p(x_k | z_{1:k}) \) that can be computed using the Bayes rule as

\[ p(x_k | z_{1:k}) = \frac{p(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})}, \]  

(29)

where

\[ p(z_k | z_{1:k-1}) = \int p(z_k | x_k) p(x_k | z_{1:k-1}) \, dx_k. \]  

(30)

The priori distribution \( p(x_k | z_{1:k-1}) \) is computed at the previous step of prediction. Using GPs regression, the function \( h_k \) can also be approximated by a normal pdf with the mean and covariance represented by Equations (2) and (3), respectively. Therefore, the distribution of \( h_k \) can be described as

\[ p(z_k | x_k) \approx \mathcal{N}(GP_{\mu}^h(x_k, D_h), GP_{\Sigma}^h(x_k, D_h)). \]

Therefore,

\[ z_k = GP_{\mu}^h(x_k, D_h) + \delta_k, \]

\[ \delta_k \sim \mathcal{N}(0, GP_{\Sigma}^h(x_k, D_h)). \]  

(31)

Generally, because of using GPs, the integral of Equation (30) approximately produces a Gaussian pdf, and therefore, the Bayes rule of Equation (29) will produce a Gaussian distribution.

Finally, merging Equations (28) and (31) results in the following equation:

\[ x_k = GP_{\mu}(x_{k-1}, u_{k-1}), D_f) + \epsilon_k, \]

\[ z_k = GP_{\mu}^h(x_k, D_h) + \delta_k, \]

\[ \epsilon_k \sim \mathcal{N}(0, GP_{\Sigma}(x_{k-1}, u_{k-1}, D_f)), \]

\[ \delta_k \sim \mathcal{N}(0, GP_{\Sigma}^h(x_k, D_h)), \]  

(32)

where \( GP_{\mu}(x_{k-1}, u_{k-1}), D_f) \) is the GP mean function of \( f \) that maps \( x_{k-1}, u_{k-1} \) to \( x_k \) by using the training data set \( D_f \), and \( GP_{\mu}^h(x_k, D_h) \) is the GP mean function of \( h \) that maps \( x_k \) to \( z_k \) by using the training data set \( D_h \). Furthermore, \( GP_{\Sigma}(x_{k-1}, u_{k-1}, D_f) \) and \( GP_{\Sigma}^h(x_k, D_h) \) are GP covariance functions of \( f \) and \( h \), respectively.

This GPs regression model will be substituted with process and measurement functions in EKF. The resulting GP-EKF is provided in Figure 1.

Algorithm GP-EKF(\( x_{k-1}, P_{k-1}, z_k \))

for \( k = 1,2,3, \ldots \) do

Prediction:

1: Calculate estimate \( \hat{x}_k \) and process noise covariance matrix \( Q_k \)

\[ [\mu_k, Q_k] = \text{GaussianProcess} \_ \text{Fun}(\hat{x}_{k-1}, D_f) \]

\[ \hat{x}_k = \hat{x}_{k-1} + \mu_k \]

2: Calculate jacobian matrix from process model

\[ [A \text{ tmp}] = \frac{\partial}{\partial x_k} \text{GaussianProcess} \_ \text{Fun}(\hat{x}_{k-1}, D_f) \]

\[ G_k = I + A \]

3: Calculate covariance matrix

\[ R_k = G_k P_{k-1} G_k^T + Q_k \]

Measurement Update

1: Calculate estimate \( \hat{z}_k \) and process noise covariance matrix \( R_k \)

\[ [\hat{z}_k, R_k] = \text{GaussianProcess} \_ \text{Fun}(\hat{x}_k, D_h) \]

2: Calculate jacobian matrix from process model

\[ [H_k \text{ tmp}] = \frac{\partial}{\partial z_k} \text{GaussianProcess} \_ \text{Fun}(\hat{x}_k, D_h) \]

3: Calculate gain matrix

\[ K_k = P_k H_k^T (H_k P_k H_k^T + R_k)^{-1} \]

4: Calculate the estimated state and covariance matrix

\[ x_k = x_k + K_k (z_k - \hat{z}_k) \]

\[ P_k = (I - K_k H_k) P_k \]

end for

Figure 1. GP-EKF algorithm.

### 3.2. Gaussian process incorporated with UKF

In this subsection, a new GP-UKF algorithm is proposed in which, instead of using the exact process and measurement models, their equivalent GP regression is used. Therefore, combining Equation (32) with the UKF is the basis for the GP-UKF algorithm, as shown in Figure 2.

The flowchart of GP-EKF and GP-UKF algorithms in fault detection is shown in Figure 3.

### 3.3. Fault detection using GP-EKF and GP-UKF algorithms

Assume that the occurrence of a fault in the system’s state equations can be modelled by

\[ x_k = f(x_{k-1}) + v_{k-1}, \]

\[ z_k = g(x_k) + w_k + L_k f_k, \]  

(33)

where \( f_k \) denotes the fault occurred at time \( k \) and \( L_k \) is a known matrix of fault. In other words, an additive fault was considered in the sensors.

One of the well-known methods in dynamic system’s fault detection is to compare the system’s output with a
faultless counterpart. Afterwards, the resulting signal (the residue signal) will be used for fault detection. In general, the procedure for fault detection can be expressed in two steps: (1) residue generation and (2) residue evaluation.

### 3.3.1. Residue generation

The first step in fault detection is to generate a signal that is sensitive to fault presence and can show the fault occurrence as soon as possible. A simple residue signal can be defined as

$$r_k = z_k - \hat{z}_k,$$  

where \( r_k \) is the generated residue, \( z_k \) is the system’s output in the faultless condition and \( \hat{z}_k \) is the estimated output.

### 3.3.2. Residue evaluation

The second step in fault detection is to define proper functions for the generated residue evaluation so that fault occurring in the system can be detected correctly. In the ideal case, if no fault occurs, the residue will be zero and otherwise, it will be non-zero. In this case, residue evaluation can be defined as

$$\begin{cases} r_k = 0 & \text{fault free}, \\ r_k \neq 0 & \text{faulty}. \end{cases}$$

But in many systems, the residue might be non-zero even though no fault has occurred; therefore, the evaluation function of Equation (35) will not be proper. For this purpose, a statistical evaluation function can be defined as

$$\begin{cases} r_k < \bar{r}_k - \lambda \sigma_r & \text{faulty}, \\ \bar{r}_k - \lambda \sigma_r \leq r \leq \bar{r}_k + \lambda \sigma_r & \text{fault free}, \\ r_k > \bar{r}_k + \lambda \sigma_r & \text{faulty}. \end{cases}$$

where

$$\bar{r} = E[r_k] = \frac{1}{m} \sum_{k=1}^{m} r_k,$$

$$\sigma_r^2 = E[(r_k - \bar{r}_k)^2] = \frac{1}{m} \sum_{k=1}^{m} (r_k - \bar{r}_k)^2$$
and the parameter \( \lambda \) should be chosen based on a trade-off between maximizing the probability of fault detection and minimizing the probability of wrong fault alarm.

Assuming that \( r \) is a GRV with the mean \( \bar{r}_k \) and covariance \( \sigma_{r_k} \), choosing \( \lambda = 1 \), \( p(\bar{r}_k - \sigma_{r_k} \leq r \leq \bar{r}_k + \sigma_{r_k}) = 0.683 \), from the probability theory. In other words, if \( \bar{r}_k - \sigma_{r_k} \leq r \leq \bar{r}_k + \sigma_{r_k} \), one is 68.3% sure that no fault has occurred. Furthermore, for \( \lambda = 2 \) and \( \lambda = 3 \), the probability with regard to the evaluation function will be

\[
\begin{align*}
&\begin{cases}
p(\bar{r}_k - 2\sigma_{r_k} \leq r \leq \bar{r}_k + 2\sigma_{r_k}) = 95.8\%, \\
p(\bar{r}_k - 3\sigma_{r_k} \leq r \leq \bar{r}_k + 3\sigma_{r_k}) = 98.4%. 
\end{cases}
\end{align*}
\]

It should be noticed that the computational burden of the proposed methods is more than that of the EKF and UKF because of the GPs regression method used in GP-EKF and GP-UKF. In other words, in the GP-EKF and GP-UKF, the system’s model is substituted with the GPs and therefore, the computational complexity of adding the GP is the difference. In addition, due to the Jacobian matrix computation in the GP-EKF algorithm, the computational burden of GP-EKF is more than that of GP-UKF (Ko & Fox, 2009).

4. Simulation results

To show the performance of the two proposed algorithms in fault detection, they were applied to an air-traffic tracking and control system. The aeroplane movement dynamics is defined by a non-linear equation as

\[
x_k = \begin{bmatrix}
1 & \sin(\omega T) & 0 & -(1 - \cos(\omega T)) & 0 \\
0 & \cos(\omega T) & 0 & \omega & 0 \\
0 & -(1 - \cos(\omega T)) & 1 & \sin(\omega T) & 0 \\
0 & \omega & 0 & \cos(\omega T) & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix} x_{k-1} + \begin{bmatrix}
\frac{1}{2}T^2 \\
T \\
0 \\
\frac{1}{2}T^2 \\
0 \\
T \\
0 \\
0 \\
0
\end{bmatrix} v_{k-1}, \tag{38}
\]

where \( T \) denotes the sampling period and \( x \) is the state vector that contains position and velocity in the \( x \)- and \( y \)-axis directions and also aeroplane’s angular speed \( (x_5 = \omega) \). The radar system used to give the measurements can also be modelled by the following equation:

\[
z_k = \begin{bmatrix}
\sqrt{x_{1_k}^2 + x_{2_k}^2} \\
\tan^{-1}\frac{x_{2_k}}{x_{1_k}}
\end{bmatrix} + w_k, \tag{39}
\]

In these simulations, 50 random points have been chosen to create the training data set. The Monte Carlo method (with 50 runs) has been used in order to provide trustful results.

To illustrate the performance of the GP-UKF, the mean and the variance of the first estimated state have been shown in Figures 4 and 5, respectively. Since other estimated states have similar results, only the first state has been shown. In Figure 4, a comparison has been made between the mean of the first estimated state and the mean of the actual state. Figure 5 shows the mean and uncertainty of the first state at each sample time.

The occurrence of three types of faults (abrupt, incipient and intermittent) was considered in the system’s output which are assumed to be additive. In these simulations, \( L_k = [11]^T \) and \( f_k \) was considered to be of the following
three different types:

\[ f_{k, \text{abrupt}} = \begin{cases} 0.125 & 20 \leq k \leq 50, \\ 0 & \text{otherwise} \end{cases}, \quad (40) \]

\[ f_{k, \text{incipient}} = \begin{cases} 0.125k & 20 \leq k \leq 40, \\ 0 & \text{otherwise} \end{cases}, \quad (41) \]

\[ f_{k, \text{intermittent}} = \begin{cases} 0.125 & 20 \leq k \leq 30, \\ 0.5 & 40 \leq k \leq 50, \\ 0 & \text{otherwise} \end{cases}. \quad (42) \]

Here, GP-EKF and GP-UKF algorithms are used for fault detection in the air-traffic tracking system. As can be seen in Figures 6–8, fault occurrence is clear in the residue signal. The ideal residue evaluation function of Equation (35) is used here. As shown in these figures, just as the fault occurs, it affects the residual signal. However, the main problem of using EKF is that it has a high computational demand and it can be used only in systems with a few number of state variables.

Figures 6–8 show the residue signal generated by GP-UKF and GP-EKF algorithms due to three fault types: abrupt, incipient and intermittent. As it is seen, the residue signals have a non-zero value when a fault occurs and they are zero in the faultless moments. The residue signal of the GP-UKF algorithm has a better quality in comparison with that of the GP-EKF algorithm since the faultless moments are clearly distinguishable from the faulty moments. In general, the value of the GP-UKF residue signal is larger than the GP-EKF counterpart in the case
of a fault occurrence, which is a major criterion for fault detection. Furthermore, the computational complexity of the GP-UKF is less than that of the GP-EKF, which makes it the proper choice for use in high-dimensional systems. The major advantage of both these algorithms is that they do not require any model of the system and the fault can be detected using only the input–output data set, while the Bayesian methods used so far for fault detection require such a model.

In another simulation, the evaluation function described in Equation (36) is used for residue evaluation in Figures 9–14, with $\lambda = 3$. Comparing Figures 9–11 with Figures 12–14 shows the superiority of the GP-UKF algorithm. In these figures, the probability of fault
occurrence equals 98%, when the residual signal is out of the given interval.

5. Conclusion

In this paper, two new methods of fault detection (GP-EKF and GP-UKF) were proposed for non-linear systems. Each of these methods was considered separately and their performance was studied using a practical highly non-linear system. The main advantage of the proposed methods in comparison with the previous methods is that they do not need an accurate model of the system while they generate accurate residue signal. Finally, a comparison was made between these two methods and the superiority of GP-UKF was shown since it generated more accurate residue signals and also had less computational volume.

References


