Fault detection and isolation in inertial navigation systems with SDRE non-linear filter

Ivan Vitanov and Nabil Aouf
Cranfield University
Shrivenham, Swindon, SN6 8LA, UK

Abstract

To ensure safe operation of unmanned aerial vehicles (UAVs), component hazards need to be dealt with robustly through effective fault diagnosis. The extended Kalman filter (EKF) has proven popular in this realm, though it is known to suffer from linearisation errors that can degrade performance. This paper presents results from initial simulation trials that address sensor fault detection and isolation (FDI) in an inertial navigation system (INS) of the type commonly mounted on smaller unmanned aircraft. We apply the state-dependent Riccati equation (SDRE) filter in this task and compare its performance to the EKF.

1. Introduction

Unmanned aircraft are set to attain ever-greater degrees of autonomy, ipso facto widening the scope of application of this constantly evolving technology. Currently, it is the defence industry that has made most progress in exploiting pilotless aircraft, though the civilian sector is catching up in its readiness to exploit these aircraft, such as in agriculture and weather monitoring. Additionally, micro UAVs are proving useful as research platforms for the development of sensors and sensor systems. With increased autonomy, the need for guidance and navigation systems that meet the highest standards of integrity and robustness is paramount.

An inertial navigation system fitted with 3-axis gyroscopes and accelerometers is an industry standard for UAV localisation, i.e., position, velocity and attitude determination. It is, however, plagued by small bias errors which grow over time as a result of recursions and integrations in calculating the vehicle state. To offset these adverse effects, sensor fusion with other navigation systems, such as global positioning systems (GPS) and magnetometers, is commonly employed to provide more accurate state estimation. However, there are environments (e.g., indoor environments) where outages of the coupled instruments can mean the INS operating stand-alone. Thus, ensuring fault-free and stable operation of this sensor system is crucial to safe and reliable adherence to a designated flight trajectory.

A sensor fault can be defined as unexpected change in a sensor signal due to degradation or damage to a sensing instrument - resulting in a corrupted read-out. Navigation systems are prone to the occurrence of faults in the normal course of operation, which can attenuate the efficacy of the navigation solution. This calls into question the fidelity of the navigation information coming in and hinders reliable localisation of the vehicle in space. Therefore, the occurrence of a fault or faults must be detected, the faulty component(s) isolated, and perhaps mitigated by reconfiguring the system controller. In this paper we examine the first two tasks as they apply to inertial navigation systems.

The fault detection and isolation (FDI) problem comes down to a binary decision at an instant in time: either there is sufficient evidence that a fault is in progress or there is not. When a fault is determined to have occurred, the task of fault isolation is to pinpoint its location. Generally speaking, FDI techniques rely on the concept of redundancy: hardware or analytical (software). Hardware redundancy compares replicas of the same signal generated by various hardware components, such as two or more sensors measuring the same quantity. Analytical redundancy, on the other hand, replaces redundant hardware implementation with a mathematical model, alongside an estimation technique that couples the model's output to that of a sensor being monitored. The analytical approach has the advantage that it typically does not require additional hardware. Thus it can be regarded as more cost-effective than the hardware redundancy approach [1]. This is an important consideration in UAV operation, since pilotless aircraft are typically more lightweight and streamlined than piloted ones, and vehicle weight and payload should preferably

be kept to a minimum. The analytical redundancy approach is, however, more of a challenge to implement, as it has to deal robustly with model inaccuracies (since no model is exact, this is critical), the presence of noise and unknown disturbances.

FDI takes a two-pronged approach to fault diagnosis - through the generation of residuals to provide diagnostic signals and use of a threshold function to distinguish faults from disturbances. The FDI filter has to be designed such that it minimises the effect of non-fault signals (model uncertainties, noise disturbances) on a residual and maximises that of the fault(s). For fault isolation, typically a set of structured residuals is used, each residual accounting for the occurrence of a fault on a single component. Alternatively, a given combination of residuals exceeding their thresholds can also provide a unique signature to flag up a fault at a specific location.

Although the Kalman filter is capable of providing real-time vehicle position estimation, and hence is capable of real-time INS residual generation, it is nevertheless based around linear system models. Consequently, it suffers from linearisation errors when dealing with nonlinear models [2]. In this paper, we investigate an alternative to EKF-based residual generation. This alternative employs State-Dependent Riccati Equations (SDRE) non-linear filtering. SDRE techniques are emerging as optimal non-linear control and filtering methods. Over recent years, various SDRE-based design approaches have been successfully applied to aerospace problems [3].

The remainder of this paper is organised as follows: Section 2 reviews inertial navigation systems used for localisation of airborne systems. Section 3 reviews the EKF and SDRE non-linear filters for FDI. Section 4 presents the results of a simulation study of INS fault diagnosis using banks of EKF and SDRE filters. Section 5 concludes the paper.

2. Inertial navigation systems

Inertial navigation is achieved by integrating the output of a set of sensors to compute position, velocity and attitude. The sensors used are a set of three gyroscopes to measure roll, pitch and yaw rotation rates (p, q, r); as well as three accelerometers measuring linear accelerations (ax, ay, az) along the three body axes, with respect to an inertial frame. Collectively the three accelerometers and three gyros make up the core sensing device of the INS, the inertial measuring unit (IMU).

The measurements from the IMU are processed through a series of integrations and transformed into an appropriate navigation frame - such as an earth-centred earth-fixed frame (ECEF) - yielding aerial position coordinates (X, Y, Z), velocities (U, V, W) and attitude Euler angles (φ, θ, ψ).
The INS can be represented in the following continuous-valued non-linear state space form:

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x,u) \end{aligned} \tag{1}$$

With x the state vector, and u the input vector to the INS (angular rates and accelerations) – alternatively this is the output vector of the IMU. That is:

$$x = [X, Y, Z, U, V, W, \varphi, \theta, \psi]^T \tag{2}$$

$$u = [p, q, r, ax, ay, az]^T \tag{3}$$

The navigation equations require that we define at minimum two reference frames. One is a vehicle coordinate frame (body or inertial); the other is a navigation frame. System equations of motion can then be derived through basic integrations and frame transformations.

By integrating the following equation we can evaluate the Euler angles (φ, θ, ψ):

$$\begin{bmatrix} \dot{\varphi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} 1 & \sin(\varphi)\tan(\theta) & \cos(\varphi)\tan(\theta) \\ 0 & \cos(\varphi) & -\sin(\varphi) \\ 0 & \sin(\varphi)\sec(\theta) & \cos(\varphi)\sec(\theta) \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} \tag{4}$$

We now have the attitude of the aircraft. Using the orientation values and the outputs of the accelerometers (*ax*, *ay*, *az*), we can then arrive at the vehicle accelerations in the body frame, given an IMU-positioned at the vehicle centre of gravity:

$$\begin{bmatrix} \dot{U} \\ \dot{V} \\ \dot{W} \end{bmatrix} = \begin{bmatrix} ax + Vr - Wq + g\sin(\theta) \\ ay - Ur + Wp - g\cos(\theta)\sin(\varphi) \\ az + Uq - Vp - g\cos(\theta)\cos(\varphi) \end{bmatrix} \quad \textbf{(5)}$$

Where g is the acceleration due to gravity. When integrated with respect to time this acceleration vector gives the body velocities (*U*, *V*, *W*) as follows:

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \int \begin{bmatrix} \dot{U} \\ \dot{V} \\ \dot{W} \end{bmatrix} dt \quad \textbf{(6)}$$

The position in the body frame can next be determined by integration of the velocity vector. If we simultaneously transform the velocity to the navigation frame, we obtain the position coordinates (*X*, *Y*, *Z*) in the navigation frame:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \int C_{bn}^{T}(\varphi,\theta,\psi) \begin{bmatrix} U \\ V \\ W \end{bmatrix} dt \quad \textbf{(7)}$$

Where C^{T}_{bn} is the transform matrix from the body frame to the navigation frame. Combining the transformation expressions in (4), (5) and (7) above into a single matrix gives us our transition function f(*x*,*u*). In our simulation experiments we assumed the observation function h(*x*,*u*) to be a unit matrix.

A difficulty with the INS is that it tends to drift as a cubic function of time due to accumulation of biases or errors [3]. The greater part of the INS errors are imputed to the inertial sensors (instrument errors).
Calibration of the INS can cancel out some of the error sources, but some residual errors will inevitably remain behind. The effects of integration compound these: errors in the accelerations and angular rates lead to an accumulation of position and velocity errors. The dominant INS error sources can be listed as follows [4]:

- Alignment errors.
- Accelerometer bias.
- Non-orthogonality of gyros & accelerometers.
- Gyro drift due to temperature change.
- Gyro scale factor error.
- Random noise.

3. Non-linear filters

We first review the linear Kalman filter (KF) by way of establishing some common notions subsequently built upon in the development of non-linear filters. The discrete linear Kalman filter tries to estimate the state x of a controlled process (in discrete time) that is governed by the following equations:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k \\ z_k &= Hx_k + v_k \end{aligned} \quad \textbf{(8)}$$

The first equation is the state equation and the second is the output equation. Between them, these two equations describe a discrete process of linear type. The equations in (8) contain the following terms: *A, B,* and *H* are matrices; *k* is the time index; *x* is called the state of the system; *u* is a known input to the system (called the control signal); *z* is the measured output; *w* and *v* are Gaussian noise terms - *w* is known as the process noise, and *v* is known as the measurement noise.

In state estimation problems, we want to estimate *x* because it contains all the necessary descriptive information about the system. The difficulty is that we cannot measure *x* directly. Instead we measure *z*, which is a function of *x* that is corrupted by noise *v*. We can use *z* to help us obtain an estimate of *x*, but we cannot typically take the information from *z* at face value because it is corrupted by noise. In a discrete-time implementation, matrix *A* relates the state at time *k+1* to the state at the previous time *k*. Matrix *B* relates the optional control input to the state *x*. Matrix *H* relates the state to the measurement.

To estimate the system state, we could just use z_k as our position estimate but for the fact that it is noisy. A more accurate estimate is provided by the KF. This is because the KF uses not only the measurement z_k, but also the information that is contained in the state equation. The KF equations can be written as follows:

$$\begin{aligned} K_k &= P_k H^T (HP_k H^T + R)^{-1} \\ \hat{x}_{k+1} &= (A\hat{x}_k + Bu_k) + K_k (z_k - H\hat{x}_k) \\ P_{k+1} &= A(I - K_k H)P_k A^T + Q \end{aligned} \tag{9}$$

Where $\hat{X}_k$ is the estimate of x_k; K_k is the Kalman gain; P_k is the estimation-error covariance matrix; Q is the covariance of the process noise, w_k, and R is the covariance of the measurement noise, v_k; I is an identity matrix. The goal of the KF is to find an a-posteriori state estimate as a linear combination of an a-priori estimate and a weighted difference between the measurement and predicted state. *K* is the gain factor that minimises the a-posteriori error covariance. The difference is called the innovation or residual, as expressed by the term:

$$(z_k - H\hat{x}_k) \tag{10}$$

In detecting faults using the KF, we record successive values of this term to generate a set of residuals – when its value peaks suddenly, this points to an abrupt change in system behaviour that could be a fault. The iterative nature of the KF results in a permanent prediction/correction cycle. To set the cycle going, we need to start with an estimate of the state at the initial time. We also need to start with an initial estimation-error covariance, P_0, which represents our uncertainty in our initial state estimate.

3.1 The extended Kalman filter

The KF is a linear filter applicable to a linear system. It no longer provides adequate results when applied to a system that does not behave linearly, even over a small range of operation. In such cases, it becomes necessary to resort to non-linear filters. Non-linear filtering can be complex. It is not as well-understood as linear filtering. However, some non-linear estimation methods have gained widespread acceptance. One of these is the extended Kalman filter, which, as its name suggests, is a non-linear extension of the Kalman filter. The EKF is predicated on the principle that if we linearise a non-linear system at certain points, then we can use linear estimation methods (such as the KF) to estimate the states. In order to linearise a non-linear system, we can use a Taylor series expansion to derive the Jacobian matrix. The EKF uses the Jacobian matrix to linearise a non-linear system model and hence derive *A* and *H* [5], [6]. The EKF algorithm can be summarised as follows [7], [8]. Non-linear system equations:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + w_k \\ z_k &= h(x_k) + v_k \end{aligned} \tag{11}$$

At each time step, we compute the following Jacobian matrices, evaluated at the current state estimate:

$$\begin{aligned} A_k &= f'(\hat{x}_k, u_k) \\ H_k &= h'(\hat{x}_k) \end{aligned} \tag{12}$$

We then evaluate the following EKF equations:

$$\begin{aligned} K_k &= P_k H_k^T (H_k P_k H_k^T + R)^{-1} \\ \hat{x}_{k+1} &= f(\hat{x}_k, u_k) + K_k (z_k - h(\hat{x}_k)) \\ P_{k+1} &= A_k (I - K_k H_k) P_k A_k^T + Q \end{aligned} \tag{13}$$

3.2 The state-dependent Riccati equation filter

State-dependent Riccati equation (SDRE) based techniques are gaining in importance as general-purpose non-linear filtering tools. One attractive feature of SDRE non-linear filters is that they overcome a number of limitations observed in classical filters, such as linearisation and initialisation issues. The SDRE non-linear filter is derived using a state-dependent coefficients (SDC) factorisation. As for the EKF (11), we model the system using a discrete-time non-linear state-space system representation. The system is assumed to be disturbed by Gaussian process noise, w_k, and measurement noise, v_k, with covariance matrices Q and R respectively:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + \Gamma w_k \\ z_k &= h(x_k) + v_k \end{aligned} \tag{14}$$

The Γ matrix is a weighting matrix for the process noise. To transform this system into SDC form we perform the following factorisation of the state and observation equations:

$$\begin{aligned} f(x_k, u_k) &= F(x_k, u_k)x_k \\ h(x_k) &= H(x_k)x_k \end{aligned} \tag{15}$$

The INS model is readily converted to SDC form by way of some mathematical transformations. This makes the application of the SDRE non-linear filter practical. The SDC form in (15) above as used in the filter equations is as follows:

$$\hat{x}_{k+1} = F(\hat{x}_k, u_k)\hat{x}_k + K_f(\hat{x}_k)(z_k - H(\hat{x}_k)\hat{x}_k) \tag{16}$$

Where:

$$K_f(\hat{x}_k) = P_k H^T(\hat{x}_k) R^{-1} \tag{17}$$

And P_k is the positive definite solution to the Riccati equation:

$$F(\hat{x}_k, u_k)P_k + P_k F^T(\hat{x}_k, u_k) - P_k H^T(\hat{x}_k) R^{-1} H(\hat{x}_k) P_k + \Gamma Q \Gamma^T) = 0 \tag{18}$$

Though the SDRE filter avoids the linearisation problem encountered with the EKF in the non-linear case, the Gaussian noise assumption and restriction on process accuracy and on the observation model are two drawbacks of this filter [3], [4].

4. Simulation results

This section presents the results of initial simulation trials using banks of six SDRE and six EKF filters (one per instrument for the combined total of six gyros and accelerometers).

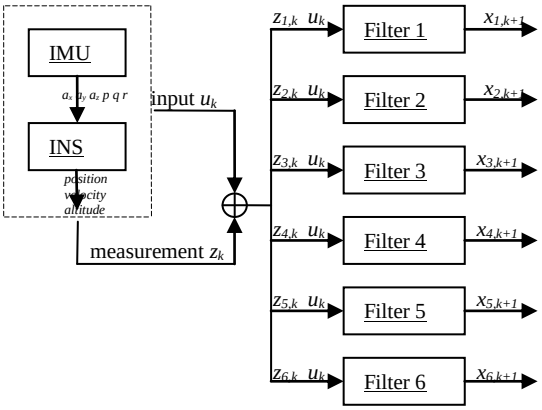

Figure 1: Bank of filters architecture

Each filter generates a residual to detect and isolate faults on the sensor coupled to it. Residuals are generated as per expression (10). In effect, the residual r(t) is the difference between the measured output y(t) and the output ŷ(t) predicted by the filter model; that is, r(t)=y(t)- ŷ(t). The filter bank architecture is depicted in Figure 1 above. Applying a residual evaluation function, consisting of static and gradient thresholds, yields the binary fault vector ε = [ε1 ε2 ε3 ε4 ε5 ε6].

4.1 EKF implementation

Figure 2 below displays the results generated using the bank of six EKF filters. Each residual is labelled according to the sensor signal handled by the corresponding filter's observation model. Perforce, the observation for each filter is a scalar value.

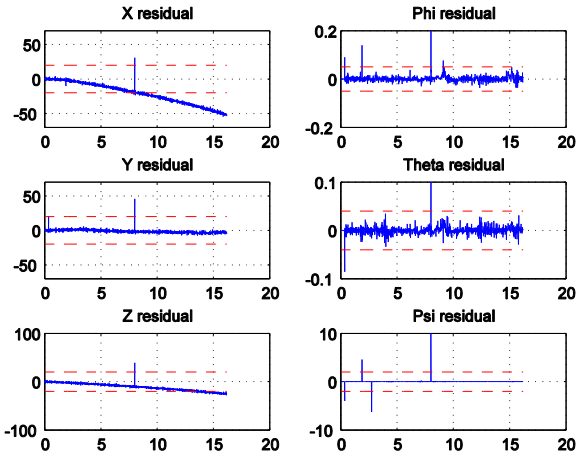

Figure 2: Fault residuals (EKF approach)

The dashed lines in Figure 2 represent the static thresholds. The case represented here is that of a simultaneous occurrence of faults on all the instruments; i.e., **ε** = [1 1 1 1 1 1] is what the fault vector should read, since the instruments are effectively decoupled. The dataset used in these trials was recorded on an actual UAV and has time duration of 1620 time steps. A continuous fault is introduced across all sensors at t=800. This can be seen to coincide with spikes at those time-points for all of the sensors, which surpass the thresholds. Spikes of large magnitude unrelated to the simulated faults (and evidenced prior to their introduction) are seen to occur on two of the attitude sensors (φ & ψ) triggering false alarms. Similarly, the x-position sensor's residual slopes downward - an occurrence unrelated to the fault incidence since it manifests even when no fault is simulated. The simulation trials with the EKF yielded a false positive rate of *5.36%* and false negative rate of *0%*.

4.2 SDRE implementation

Figure 3 overleaf shows results generated using a bank of SDRE filters, with the same fault profile as in the EKF experiment. Although somewhat more oscillatory, the SDRE residuals stay within their bounds until the faults are introduced, with the exception of the extraneous spikes occurring on the residuals for attitude sensors φ & ψ, as was the case with the EKF. Hence, there are no sloping residuals due to deteriorating estimation quality, so the signal represents an enhancement over the EKF.

A further advantage regarding fault detection capability issues from the fact of the gentler gradient of the spiking residual in the fault region. Whereas the EKF produces sharp vertical spikes (after a fault), which are similar, gradient-wise, to those that are spurious, in the case of the SDRE residuals, the spurious and non-spurious peaks are more markedly differentiated, aiding residual evaluation.

In addition, some of the SDRE residuals demonstrate somewhat of a step-change response in the event of a fault occurring. This can be advantageous over the characteristic vertical fault spikes of the EKF residuals, where the fault is quickly damped out of the estimation, because it provides a more characteristic fault signature and there is more of a lingering effect of the fault on the residual.

The false positive and negative rates recorded are *1.51%* and *3.85%* respectively. It is noteworthy that the latter would be close to zero, but for an occasional drift in the position residuals, causing a dip below the negative threshold before the fault occurs – the fault then spikes up into the neutral band between the thresholds leading to misdetection.

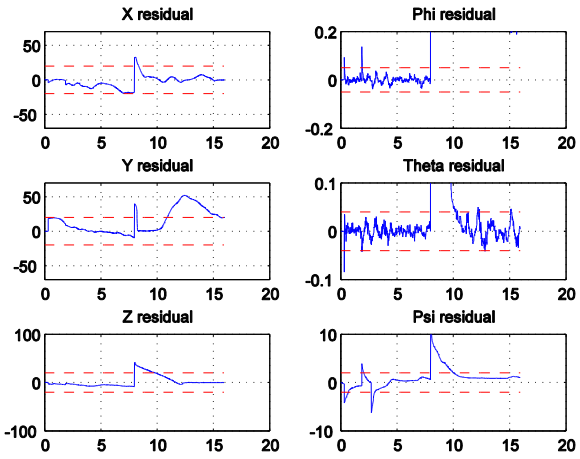

Figure 3: Fault residuals (SDRE approach)

5. Conclusion

In this paper, we have proposed a non-linear filter architecture to perform fault detection and isolation on the sensor gyros and accelerometers of a strap-down INS (as would be found on-board a small UAV) subject to Gaussian noise.

We have compared two approaches using this framework through filter banks of SDRE and EKF filters. Initial simulation results are encouraging and demonstrate that the SDRE filter is indeed fit for the purpose of FDI in an aerial navigation setting in the presence of Gaussian noise. In our future work we aim to build on these results to improve robustness to disturbances, modelling inaccuracies, and noise, by elaborating on the SDC linearisation principle.

References

[1] Hwang, I., S. Kim, Y. Kim and C. E. Seah. 2010. A survey of fault detection, isolation, and reconfiguration methods. *IEEE Trans. On Control Syst. Technology*. Vol. 18, no. 3.

[2] Simon, D. Optimal state estimation, Kalman, H∞ and nonlinear approaches. 2006. *Wiley*. Pp. 333–357.

[3] Nemra, A. 2010. Robust airborne 3D visual simultaneous localisation and mapping. PhD Thesis. Cranfield University.

[4] Nemra, A., and N. Aouf. 2010. Robust INS/GPS sensor fusion for UAV localization using SDRE nonlinear filtering. *IEEE Sensors Journal*. Vol. 10, No. 4, April

[5] Einicke, G., and L. White. 1999. Robust extended Kalman filtering. *IEEE Transactions on Signal Processing*. Vol. 47, No. 9, pp. 2596 - 2599.

[6] Shaked, U., and N. Berman. 1995. H∞ Nonlinear filtering of discrete-time process. *IEEE Trans. Signal Processing*. Vol. 43, pp. 2205-2209.

[7] Vikas Kumar N. 2004. Integration of inertial navigation system and global positioning system using Kalman filtering. MTech Dissertation. Department of Aerospace Engineering, Indian Institute of Technology, Mumbai.

[8] Schmidt, G.T. 1978. Strapdown inertial systems - theory and applications. AGARD Lecture Series, No. 95.