We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

A Reference Recursive Recipe for Tuning the Statistics of the Kalman Filter

Mudambi R Ananthasayanam

Additional information is available at the end of the chapter

Abstract

The philosophy and the historical development of Kalman filter from ancient times to the present is followed by the connection between randomness, probability, statistics, random process, estimation theory, and the Kalman filter. A brief derivation of the filter is followed by its appreciation, aesthetics, beauty, truth, perspectives, competence, and variants. The menacing and notorious problem of specifying the filter initial state, measurement, and process noise covariances and the unknown parameters remains in the filter even after more than five decades of enormous applications in science and technology. Manual approaches are not general and the adaptive ones are difficult. The proposed reference recursive recipe (RRR) is simple and general. The initial state covariance is the probability matching prior between the Frequentist approach via optimization and the Bayesian filtering. The filter updates the above statistics after every pass through the data to reach statistical equilibrium within a few passes without any optimization. Further many proposed cost functions help to compare the present and earlier approaches. The efficacy of the present RRR is demonstrated by its application to a simulated spring, mass, and damper system and a real airplane flight data having a larger number of unknown parameters and statistics.

Keywords: adaptive EKF, expectation maximisation, maximum likelihood, Cramer Rao bound, probability matching prior

1. Brief introduction to the historical development of Kalman filter (KF)

1.1. Conceptual beginning of KF in ancient Indian astronomy

As is well known if there is one thing that does not change in nature it is the change. Such a change has to be captured by some means. In general, neither the change nor the capture is exact. Hence based on some suitable criteria a combination can be derived to correct. Such an

update process has to be repeated at suitable intervals. The above philosophy of 'change, capture, and correct' is the one that is followed in the Kalman filter. The ancient Indian astronomers had understood the above philosophy.

The ancient Indian astronomers, at least since AD 500, used the above concept to update the parameters for predicting the position of celestial objects for timing their Vedic rituals based on measurements carried out at various time intervals which can be stated as

$$\text{Updated parameter} = \text{Earlier parameter} + (\text{Some quantity})$$
$$\times (\text{Measured} - \text{Predicted}) \text{ Position of the celestial object} \tag{1}$$

The 'some quantity' as we will see later on is the Kalman gain. Further, note the measured longitude of the celestial object is different from the state that is updated, which is the number of revolutions in a yuga just as state and measurements are in general different in many Kalman filter applications!

They needed to calculate the position of the celestial objects like Sun, Moon, and other planets to carry out the Vedic rituals. But their predicted positions changed over many centuries due to unmodeled or unmodelable causes. The French historian Billard [1] noted that measurements were carried out (in fact extending over many years or even decades!) at various times starting from around AD 500 by Aryabhata to AD 1600 and the parameters were corrected to make the predicted position of the objects consistent with new observations. **Table 1** shows such revisions over a period of time from Sarma [2].

During the above period, Nilakantha (around AD 1443) had stated that the eclipses cited in Siddhantas can be computed and the details verified. Similarly, other known eclipses, as well as those currently observable, are to be studied. In the light of such experience, future ones can be computed and predicted (**extrapolation!**). Or eclipses occurring at other places can be studied taking into account the latitude and longitude of the places and on this basis the method for the true Sun, Moon… can be perfected **(data fusion!).** Based on these, the past and future eclipses of one's own place can be studied and verified with appropriate refinement of the technique. This is just the idea of **'smoothing'!** Billard had a problem for later Indian

Source, Number of days per yuga +	Date AD	Moon 5,77,53 +	Moon's Apogee 4,88,000 +	Moon's Node 2,32,000 +	Mars 22,96,000 +	Mercury 1,79,37,000 +	Jupiter 3,64,000 +	Venus 70,22,000 +	Saturn 1,46,000 +
Aryabhatiyam, 500	499	336	219	226	824	020	224	388	564
Parahita,500	563	314.2	121.9	307.3	862.3	378	184.0	257.8	581.0
Drgganita,517.019	1431	321.009	123.229	297.832	862.959	072.112	172.296	270.775	626.695
Tantrasangraha 517.019	1443	320	122	300	864	058	180	268	612
Golasara,517.019	1486	320	120	300	864	040	180	270	600
Siddhantadarpana, 839.500	1500	332.321	123.318	226.745	862.137	120.175	160.611	270.552	571.016
Drkkarana,839.500	1607	320	122	298	863	100	166	272	612

Table 1. Corrections of planetary parameters in ancient India. From Sarma [2] (Sun 43,20,000; Number of days per yuga = 1,57,79,17,000+).

astronomers as 'If the elements of Aryabhata are now wrong, they must have been accurate when he was living. Thus, we ought to **establish the astronomical elements upon both on his own at his time and the new observations of the present time**'. He noted that based on the above reasoning, some Indian canons were evolved and one such canon around AD 898 shows a very high accuracy valid over a larger number of centuries [1].

1.2. Development of the concepts in KF during the medieval period

On January 1, 1801, Piazzi, while searching for a wrong entry in a star catalogue, discovered the largest and the first of the asteroids Ceres. He tracked its position for the next 41 days, before it was lost in space. He wrote to Bode who felt it to be the missing planet in his formula. Its orbital elements could not be determined by the then available methods. Newton had stated it as the most difficult nonlinear problem then in astronomy. Piazzi's discovery was published in 1801. Gauss tackled the problem and estimated its orbit and sent it to Piazzi who found it again on the last day of 1801! Gauss published his orbit determination methods only in 1809 [3] describing his 1795 method of least squares (LS) used in his estimates of the Ceres orbit. Gauss had an ideal situation with a good system model and only the measurement noise, and thus with his least squares approach he could get an estimate and a qualitative measure for the uncertainty. Independently, the method of LS had been discovered and published by Legendre of France and Robert Adrian of the United States. In fact, even before Gauss was born, the physicist Lambert had used it! Gauss has provided almost all the essentials of present day estimation theory. He postulated that a system model should be available, minimum number of measurements for observability, the redundant data helping to reduce the influence of measurement errors, a cost function based on the difference between the measurement and that predicted by the model should be minimised. There should be some a priori knowledge concerning the unknowns to be estimated. Further, since the errors could be unknown or unknowable, he had given hints about probabilistic approach, normal distribution, and method of maximum likelihood estimation, linearization, and the Gaussian elimination procedure. Gauss did not balance the governing differential equation, but tried to fit the measurement with the prediction. If he had tried the former, he would have been led to a biased solution! Fortune favours the brave! This is where a proper mathematical framework helps to understand if an algorithm converges to the correct value with more and more measurements. The post Gaussian contributions in ET consists of the method of moments, method of maximum likelihood estimation, the Kalman filter and its variants, frequency domain approach, and handling time varying dynamical state and parameters. Further, the use of matrix theory, sequential instead of batch processing, and real-time processing by computers exist. We are dealing with more difficult situations, but the conceptual framework to solve these problems had been fully laid out by Gauss!

1.3. Mature KF during the twentieth century and beyond

Almost every concept in present day science and technology seem to have its root in ancient times as mentioned earlier. It must be understood that the classic papers by Kalman [4] and Kalman Bucy [5] was conceptually preceded earlier by Thiele as mentioned by Lauritzen [6], Swerling [7], Stratonovich [8] and some other researchers. The development of the Kalman filter is one of the most interesting and useful innovations of the twentieth century, and it owes its

origin to the least squares solution proposed by Gauss. For Gauss, the state equations were the exact Newton's laws of motion and only measurement noise existed. The sequential least squares were rediscovered by Plackett [9] and Kalman [4]. Thus, in fact, Sprott [10] has questioned if the Kalman Filter is really a significant contribution when Gauss was far ahead! The point is that the frequency domain approach of the Wiener filter [11] has been improved to the natural time domain approach. Further the shift from batch to the sequential approach is very convenient to handle continuous measurement data flow. It is to the credit of Kalman apart from unifying earlier results that he introduced the concept of controllability and observability, which means the system to be identified has to be properly excited and observed thus making the estimation problem systematic and consistent. The only slight difference, but very momentous between the Recursive Least Squares (RLS) and the Kalman filter is the time propagation of the state and covariance estimates between measurements (see for example [12]). Presently, the scale and magnitude of many difficult and interesting problems that estimation theory (ET) is handling could not have been comprehended by Gauss or Kalman. Examples are airplane flight test data analysis [13, 14], target tracking [15], evolution of the space debris scenario [16], fusion of GPS and INS data [17], study of the tectonic plate movements [18], high energy physics [19], agriculture, biology, and medicine [20], dendroclimatology [21], finance [22], source separation problem in telecommunications biomedicine, audio, speech and in particular astrophysics [23], and atmospheric data assimilation for weather prediction [24].

2. Randomness, probability, statistics, random process, estimation theory, and Kalman filter

The connection among probability, statistics, random process, and estimation theory by Ananthasayanam [12] as shown in **Figure 1**, are ubiquitous in science, technology, and life.
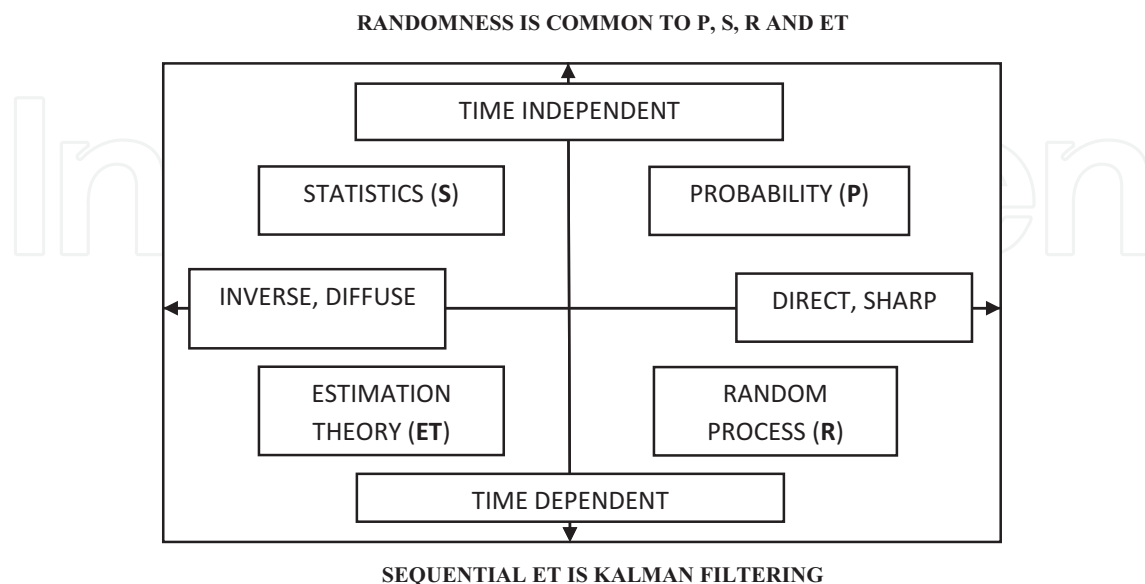
RANDOMNESS IS COMMON TO P, S, R AND ET

SEQUENTIAL ET IS KALMAN FILTERING

Figure 1. Relationship between probability, statistics, random process, and estimation theory.

The randomness is common to all and sequential time-dependent statistical analysis can be called as Kalman Filtering.

2.1. Randomness

We have moved unambiguously from determinism to randomness ever since ancient times to the present. We are compelled to understand, model, estimate and control nature which is random in a probabilistic way. Randomness occurs inevitably in all walks of life. The ontology (the true nature) is one thing and the epistemology (our understanding) is another thing. A computer generating a sequence of random numbers is deterministic ontology, but for the user who does not know how they are generated it is probabilistic epistemology. Randomness is patternless but not propertyless. Randomness could be our ignorance. Chance is no longer something to worry about or an expression of ignorance. On the contrary, it is the most logical way to present our knowledge. We are able to come to terms with uncertainty, to recognise its existence, to measure it and to show that advancement of knowledge and suitable action in the face of uncertainty are possible and rational. Random chance may be the antithesis of all law. We look for the alternatives and provide the probabilities of their happening as measures of their uncertainties. Knowing the consequences of each event and the probability of its happening, decision making under uncertainty can be reduced to an exercise in deductive logic. It is no longer a hit and miss affair. But, the way out is to discover the laws of chance and convert chance to choice in life. The aim of life is to make the earth a happy place to live.

Randomness could occur due to the uncertainty, variability, complexity, or enormity. A classic example is the deterministic coin tossing. There is enormity in dealing with a large number of air molecules, the complex interaction of air among themselves and with the coin, variability of the initial condition, and the uncertainty due to air currents. Of course, many deterministic mathematical problems could be handled using probabilistic approaches called Monte Carlo techniques. Quantum Mechanics seems to possess true randomness. One feels that randomness is a nuisance, and should be avoided. However we have to live with it and compulsively need it in many situations. In a multiple choice question paper, no examiner would dare to put all the correct answers at the same place! As another example the density, pressure, and temperature, or even many trace constituents in air can be measured with a confidence only due to the random mixing that invariably takes place over a suitable space and time scale. As we will see later, the introduction of random process noise into the kinematic or dynamical equations of motion of aircraft, missiles, launch vehicles, and satellite system helps to inhibit the onset of Kalman filter instability and thus track these vehicles. The well-known statistician Rao [25] states that statistics, as a method of learning from experience and decision making under uncertainty, must have been practiced from the beginning of mankind. The inductive reasoning in these processes has never been codified due to the uncertain nature of the conclusions drawn from any data. The break through occurred only at the beginning of the twentieth century with the realisation that inductive reasoning can be made precise by specifying additionally just the amount of uncertainty in the conclusions. This helped to work out an optimum course of action involving minimum risk in uncertain situations in the deductive process. This is what Rao [25] has stated as a new paradigm of thinking to handle uncertain situations as

$$\text{Uncertain knowledge} + \text{Knowledge of the amount of uncertainty in it}$$
$$= \text{Usable knowledge} \tag{2}$$

The Kalman filter dealing with only the estimates and their uncertainties thus has minimal additional information to handle.

2.2. General remarks on P, S, R, and ET

Though randomness exists in all the above, due to the undercurrent of some determinism wrapped by uncertainty the processes are not completely haphazard. The **P** and **S** are time independent, whereas **R** and **ET** are their time-dependent analogues. The **P** and **R** are direct using deductive logic; whereas **S** and **ET** are inverse problems using inductive logic, thus more difficult than the former. The **S** and **ET** deal with data that are insentient and do not speak. The analyst gives life to the data to find out the underlying model mechanism based on intuitive and subjective analysis to obtain results and conclusions which are meaningful and useful.

In an inverse problem with a limited sample size from the population with or without noise, the problem is considered, as well posed if (i) there exists a solution to the problem (existence), (ii) there is at least one solution to the problem (uniqueness), and (iii) the solution depends continuously on the data (stability). Generally, an acceptable and reasonable solution can be worked out by specifying some subjective criterion cast in terms of a quantitative cost function to match (in some reasonable way) the model and the measurements made on the system. Further, no matter whatever technique one adopts unless and until the model structure of the system is appropriate and the parameters in it are identified along with the noise sequence, one can never obtain the true value but only be around it all the time. When a system model characterises many effects, then the model structure should reflect each one appropriately and the parameters in them estimated for apportioning the different effects.

2.3. Probability (P)

The probabilistic approach subsumes in general the deterministic laws of physics (but one can include them also) or others postulated for any system. It utilises the axiomatic rules governing probability leading to outcomes regarding the behaviour of the ensemble. The probability itself can be broadly specified based on Classical, Bayesian, and the Frequentist approaches [26]. These are, respectively, based on the 'principle of indifference thus equiprobable', 'degree of belief', and 'limiting frequency' of the occurrence of the events. The law of large numbers and the central limit theorem under very general conditions (leading to a **Gaussian** distribution) assist in reaching practical conclusions.

2.4. Statistics (S)

In order to define statistics, it seems best to follow Feller [27] which in our language is to analyse the **measurement** data to develop a mathematical **model** with the intuitive **mind** providing the methodologies. The mind is nebulous, measurement is spotty, and model is smooth if one may say so. **Then, the simplest definition of statistics seems to be a good translation of measurements into a model by the mind.** Of course, the three are randomly

used for progress in science, technology, and life. An intuitive subjective idea is cast in an objective form to help the analyst decide if the difference between measurement and model can be attributed to statistical fluctuation or deterministic change. Such a decision is always based on extra statistical (!) implications. All the concepts and problems of statistics go over to estimation theory, and in turn to Kalman filter which is to sequentially process the time-dependent data. Generally, in statistics, **linear** relationship among the variables is prolific since it is the simplest and anything else would have to be justified *a priori*.

The philosophical discussions about the Classical, Bayesian, and Frequentist approaches to probability and statistics can go on endlessly. The most prevalent view is Frequentists deal with data only, but Bayesians try to incorporate well established '*a priori*' information as well into the problem. But no matter whichever approach is followed, eventually one has to take a practical view and ensure the final results are credible and useful.

2.5. Random process (R)

Here, in **R**, the simplest characterisation of time-dependent noise is white in time following a Gaussian distribution, thus containing the least amount of information! However, due to the undercurrent of some amount of determinism wrapped by uncertainty, the processes are not completely haphazard. If the atmosphere and earth quakes behave like white noise then meteo-rologists, geologists would have no work! The underlying deterministic processes are wrapped in noise providing some underlying correlations. The white noise is the worst data that will fail any algorithm for prediction, but used most prolifically in **ET** since it is mathematically tractable!

2.6. Estimation theory (ET)

The basic framework of **ET** in analysing a given measurement data consists of qualitatively modelling the system, measurement and all the noise characteristics, a criterion to match or mix the model output with the measurements in some optimal sense, a numerical algorithm for the above task and consequently estimate the unknown parameters and the noise statistics together with their uncertainties and lastly an internal consistency check to ensure that the assumptions regarding model and measurement noise above are consistent and if not the above steps have to be modified and repeated.

There is a general feeling that ET solutions have to be objective with little scope for subjectivity. However, it is interesting to note that subjectivity cannot be avoided, but it is the one that helps all the way from the formulation of the problem to obtaining the final solutions. Deterministic approaches of the early period have given way to probabilistic approach that is neither a fashion nor the truth (!) The probabilistic rules aid in modelling the scenario and the statistical approach of analysing the data with all its subjectivity provides acceptable quantitative esti-mates together with their uncertainties.

3. Overview of Kalman filter

The simplest formulation of a Kalman filter is when the state and measurement equations are linear, a well-known fact as mentioned by Brown and Hwang [28]. For linear systems during

evolution and update the form of the normal distribution is maintained, but with changed mean and covariance. However, the Kalman filter has found its greatest application for nonlinear systems. Particularly in many aerospace applications, the unknown parameters multiply the state variables to provide force or moments acting on the vehicles. If the unknown parameters are treated as additional states then the system of equations become nonlinear. Thus, the Extended Kalman filter (EKF) formalism can be used. The EKF formulation provides the simplest scenario to present the proposed Recurrence Recursive Recipe (**RRR**). Other filter formulations contain the effect of further approximations, discretizations and other features. A typical continuous state with discrete measurements in time forms a nonlinear filtering problem and can be written as

$$\mathbf{x}(k) = \mathbf{f}(\mathbf{x}(k-1), \mathbf{\Theta}, \mathbf{u}(k-1)) + \mathbf{w}(k) \tag{3}$$

$$\mathbf{Z}(k) = \mathbf{h}(x(k), \mathbf{\Theta}) + \mathbf{v}(k), \quad k = 1, 2, 3, \ldots.\mathbf{N} \tag{4}$$

where '**x**' is the state vector of size (n × 1), '**u**' is the control input, **Θ** is the parameter vector of size (p × 1) and '**Z**' is the measurement vector of size (m × 1). The '**f**' and '**h**' are nonlinear functions of state and measurement equations, respectively. The process and measurement noise are assumed to be zero mean with covariance **Q** and **R**, respectively, and their sequences are uncorrelated with each other. The states are generally not directly observable, but the measurements are related to the states.

In EKF formulation, the parameter vector **Θ** is augmented as additional states. Thus,

$$\begin{bmatrix} x(k) \\ \Theta(k) \end{bmatrix} = \begin{bmatrix} f(x(k-1), & \Theta(k-1), u(k-1)) \\ & \Theta(k-1) \end{bmatrix} + \begin{bmatrix} w(k) \\ 0 \end{bmatrix} \tag{5}$$

The nonlinear filtering problem is now redefined as

$$\mathbf{X}(k) = \mathbf{f}(\mathbf{X}(k-1)) + \mathbf{w}(k) \tag{6}$$

$$\mathbf{Z}(k) = \mathbf{h}(\mathbf{X}(k)) + \mathbf{v}(k), k = 1, 2, \ldots.\mathbf{N} \tag{7}$$

where '**X**' and '**w**' are, respectively, the augmented state and process noise vector of size ((n + p) × 1). The control symbol '**u**' is not shown for brevity. The solution for the above filtering problem can be summarised as

$$\text{Initial State Estimate } \mathbf{X}(0|0) = \mathbf{X0} = E\left[\mathbf{X}(t0)\right], \tag{8}$$

$$\text{Initial State Covariance Matrix } \mathbf{P}(0|0) = \mathbf{P0} = E\left[(X0 - X(t0))(X0 - X(t0))^T\right] \tag{9}$$

$$\text{Prediction Step}: \mathbf{X}(k|k-1) = \mathbf{f}(\mathbf{X}(k-1|k-1)), \tag{10}$$

$$\mathbf{P}(k|k-1) = F(k-1)P(k-1|k-1)F(k-1)^T + Q_k \tag{11}$$

We presume that $\mathbf{X}(k|k-1)$ and $\mathbf{P}(k|k-1)$ represent the estimates of the state and its covariance matrix at time index k, based on all information available up to and including time index

k−1. Then, we update the state value from $\mathbf{X}$ (k|k − 1) to $\mathbf{X}$ (k|k) using the measurement $\mathbf{Z}$(k) with uncertainty denoted by $\mathbf{R}$(k) based on the value of $\mathbf{K}$(k) called the Kalman gain such that the updated covariance $\mathbf{P}$(k|k) has the individual terms along its major diagonal is a minimum leading to

$$\text{Update Step}: \mathbf{K}(k) = \boldsymbol{P}(k|k-1)\,\boldsymbol{H}(k)^T\,[\boldsymbol{H}(k)\,\boldsymbol{P}(k|k-1)\,\boldsymbol{H}(k)T + \boldsymbol{R}(k)]^{-1} \tag{12}$$

$$\mathbf{X}(k|k) = \mathbf{X}(k|k-1) + \boldsymbol{K}(k)\,[\mathbf{Z}(k) - \mathbf{h}(\mathbf{X}(k|k-1))] = \boldsymbol{K}(k)\mathbf{v_k} \tag{13}$$

$$\mathbf{P}(k|k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]\mathbf{P}(k|k-1) \tag{14}$$

where $\mathbf{P}$ denotes the uncertainty, $\mathbf{F}(k-1)$ is the state Jacobian matrix $(\partial \mathbf{f}/\partial \mathbf{X})$ evaluated at $\mathbf{X} = \mathbf{X}(k-1|k-1)$ and the measurement Jacobian $\mathbf{H}(k)$ evaluated at $\mathbf{X} = \mathbf{X}(k)$. The $\mathbf{X}(k|k-1)$ denotes the value at t(k) based on the process dynamics between t(k-1) and t(k), but before using the measurement information. The observation (measurement) of the process is at discrete time points in accordance with the local linearised relationship $\mathbf{H} = (\partial \mathbf{h}/\partial \mathbf{X})$ evaluated at $\mathbf{X} = \mathbf{X}(k)$ at the measurement time point. The quantity which is the difference between the actual measurement and the predicted model output

$$\mathbf{v_k} = [\mathbf{Z}(k) - \mathbf{h}(\mathbf{X}(k|k-1))] \tag{15}$$

is called the innovation. Further when the measurement is compared with the updated state $\mathbf{X}$(k|k), then the quantity

$$\mathbf{v_f} = [\mathbf{Z}(k) - \mathbf{h}(\mathbf{X}(k|k))] \tag{16}$$

is called the filter residue. As the filter passes through the measurement data, the last measurement provides the best estimate using all the data points. In order to obtain similar estimates at all the intermediate time points using all the measurements, the filter can be operated backwards and with a proper blending provides the smoothed estimates such as by Rauch et al. [29]. The quantity

$$\mathbf{v_s} = [\mathbf{Z}(k) - \mathbf{h}(\mathbf{X}(k|\mathbf{N}))] \tag{17}$$

is called as the smoothed residue where $X(k|\mathbf{N})$ is the smoothed state at time t(k) based on all the measurements $\mathbf{N}$.

It may be noted that when the innovation is white it means all the information has been extracted from the data and no further information is left, with both the models and the algorithm have done their best job!

Note that we have combined the local state estimate and the measurement both at time t(k) to obtain an updated state. Further, the use of only the estimate and covariance all over the filter tacitly implies the state and measurement variables are all distributed or approximated as quasi Gaussian. Thus, if initially the state is assumed to be distributed normally with mean $\mathbf{X0}$ and covariance $\mathbf{P0}$ then the KF involves iterative operation of two steps: prediction and after an update (also called correction) step there is a subtle reset as a Gaussian. Thus, with all such subjective features, the final answer can only be an answer rather than a unique answer.

There are five steps in the Kalman filter, namely state and covariance propagation with time, Kalman gain calculation and the state and covariance updates by incorporating the measurement. The state propagation and update refer to the sample while the covariance propagation, update and the Kalman gain refer to the ensemble characteristics. It is possible that at times based on the measurements the updates may move locally away from the true values, but with increasing data it will move towards the true value. But this should be noted as the behaviour of the sample. What the filter gain denotes is that in an overall probabilistic **ensemble** sense the Kalman filter will outperform many other estimators. This is analogous in life to some righteous persons who appear to loose but that in the long run they will win! These steps that statistically combine two estimates at any given time point, one from state and the other from measurement equation, are formal if only their uncertainties denoted by their covariances are available. Thus, the states can be estimated given the initial **X0** and **P0** as well **Q** and **R** over time. Over a time span in order to match and minimise the difference called the innovation, in some best possible sense, a well-known criterion is the method of maximum likelihood estimation (MMLE). The innovation follows a white Gaussian distribution which is operationally equivalent to minimising the cost function.

$$\mathbf{J} = (\mathbf{1}/\mathbf{N})\Sigma \, \boldsymbol{\nu}_k \, [H(k)P(k|k-\mathbf{1})H(k)T + R(k)]^{-1} \, \boldsymbol{\nu}_k^T = \mathbf{J} \, (\, \mathbf{X0}, \mathbf{P0}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\Theta} \,) \qquad (18)$$

based on summation over all the **N** measurements and has to be solved for **X0, P0, Q, R, Θ.** The importance of the innovation following white Gaussian for filter performance was brought out by Kailath [30]. Generally, mathematical treatment is terse (as the original paper of Kalman deriving the filter from the orthogonal projection principle) and refers to large data. But the sensitivity of the final results to the intermediate statistical quantities for filter consistency (at what confidence level?) is not apparent and sometimes lead to physically unacceptable results as noted in Shyam et al. [31]. Hence, for engineering applications, it is desirable to look for other statistics from the filter. When **Q** ≡ 0, the MMLE is called as the output error method with the Kalman gain matrix being zero. When **Q** > 0, the method is called as filter error method. In the usual Kalman filter implementation generally one does not solve for the statistics **P0, Q** and **R** but they are adjusted manually to obtain acceptable results. The numerical effort of minimising **J** has to appear in the estimation of the filter statistics. The Kalman filter is not a panacea to obtain better results when compared to simpler techniques of data analysis. The accuracy of the results using Kalman filter depends on its design based on the choice of **X0, P0, Θ, R** and **Q**. If the above values are not chosen properly then the filter results can be inferior to those from simpler techniques.

Other cost functions can and have been used in Kalman filter work such as the Integral of Time multiplied by Absolute Error with time as a scaling factor. This is meaningful since it is important to ensure a zero error after the filter has converged. This is given by

$$\mathbf{J}_{\mathbf{ITAE}} = (\mathbf{1}/\mathbf{T}) \int (1/\mathbf{N}) \sum \mathbf{a_i \nu_i} \, dt \qquad (19)$$

where the $\mathbf{a_i}$ is suitable weight related to the innovation covariance. Another cost function useful to study the effects of inadequate modelling in state estimation problem that is very common in Kalman filter studies has been proposed and used in rendezvous and docking problem by Philip and Ananthasayanam [32] as

$$\mathbf{J_{MODEL}} = (\mathbf{1}/\mathbf{N}) \sum \big(\boldsymbol{x}(i) - \boldsymbol{x}_{ref}(i)\big) \mathbf{P}^{-1}\big(\boldsymbol{x}(i) - \boldsymbol{x}_{ref}\big)^T \tag{20}$$

with the summation is over all the **N** time points and the suffix 'ref' refers to a desired reference trajectory to be followed and the argument in x (.) denotes the time step or point. The **P** is the covariance matrix obtained with nominal values for the unknown disturbances. If the variations or a deficiency in the modelling is beyond the statistical fluctuations as denoted by the covariance, then the above cost function changes substantially and indicates a degradation of the filter performance. Further cost functions are introduced here to obtain confidence in the results from the proposed **RRR** for tuning the filter statistics.

3.1. An appreciation of the Kalman filter

Science and technology has progressed through theories and experiments. It is only in Kalman filter that both theory and experiment are handled simultaneously almost all the time. Due to the seemingly unpretentious fact of splitting the state and measurement equations and switching between the state propagation and its update using the measurements, very interesting outcomes have been shown to be possible. This is similar to any amount of deep study and understanding of the state or the measurement equations (theory and experiment) separately may not be able to comprehend the exciting possibilities and abilities when both are combined together. This is similar to the components of a watch, or the cells in an organism leading, respectively, to the time keeping ability or life, which do not exist in the individual components. The GPS is another brilliant example of such a synergism. The competence of the Kalman filter is similar to the saying 'wholes are more than the sum of their parts' as stated by Minsky [33]. It is the above feature that can be called as synergistic, parallel, operator splitting that is the remarkable and profound aspect of the Kalman filter rather than describing it as a sequential least square estimator, or capable of handling time varying states and measurements.

3.2. Aesthetics, beauty, and truth of the Kalman filter

The aesthetics of the Kalman filter is to consider only the estimate and the covariance representing the uncertainty. Just only one additional quantity to move from a deterministic case to probabilistic scenario for describing the results is economical.

In Kalman Filter, at the initial time, the probability density function is assumed to be Gaussian. This need not be true if the state is not a measured quantity whence it could be true or an assumption is made. If the state equations are nonlinear then after the state propagation the probability density function becomes non-Gaussian. Next, in general, the combination of the propagated states and measurements (with any one being non-Gaussian) need not lead to an updated Gaussian. However, at a measurement update, only the estimate and the covariance of the density functions which could be non-Gaussian is used. Thus, even in nonlinear problems after an update when the distribution need not be Gaussian is subtly reset to be a Gaussian with the updated estimate and covariance.

These above are in some ways similar to many other problems in science and engineering wherein only the first and second derivatives or moments alone are considered. This is just the

reason and fortuitously as well for using velocity and pressure or temperature in equilibrium thermodynamics, which depend on the first and second moments, respectively, of the distribution function governing the random velocity of the gas molecules. This leads to the consideration of fewer moments or states to describe the dynamics of the gas flow. What would happen if higher order moments had been relevant! Even in classical dynamics in the equations of motion, the linear and angular accelerations are only the second derivatives! As another example, take a rectangular distribution and with increase in sample size the lower order moments converge faster than the higher order moments. This is because away from the middle, the tail controls the higher order moments. For a very similar reason, the Boltzmann equation in kinetic theory deals only with single particle distribution as against multi particle distribution function.

The beauty in the Kalman filter is whether it is true or otherwise many random variables are assumed to follow a multivariate Gaussian distribution, and thus the derived joint and marginal density functions are all Gaussian. The Gaussian distribution provides an enormous amount of mathematical tractability exactly for linear systems and approximately for nonlinear systems.

The truth in the Kalman filter equations is that once it is derived in one way, it is possible to derive it in a variety of ways with slightly different assumptions, but mostly leading to similar set of basic equations as for linear problems. The author of each book has his own derivation! It is interesting to note that the simplest formulation of the Kalman filter is based on minimum amount of *a priori* knowledge or information in probability, statistics, and random process providing, respectively, the Gaussian distribution, linear relationship among the variables, and white noise [12]. Otherwise one has to justify them which may not be simple or easy. If necessary other suitable distributions, nonlinearity, and coloured noise can be introduced later into the filter framework.

3.3. Different perspectives and competence of Kalman filter

The Kalman filter can be viewed as an inverse problem, deterministic or probabilistic approach, reversing an 'irreversible' process. Also it can be considered as qualitative modelling and quantitative estimation, stochastic corrective process by Narasimha [34], data fusion and statistical estimation by probabilistic mixing, optimization in the time (or frequency) domain. The competence of the filter consists of estimating unknown or inaccurately known parameters (including deterministic errors) in the state and measurements, estimation of process and measurement noise, improved states and measurements by smoothed filter estimates, estimation of unmodelable inputs by modelling them in a probabilistic way, unobservables from observables, expansion of the scenario, handling computational errors by noise addition, consistency check of the whole process of modelling, convergence of the numerical algorithm, and the extraction of the complete information from the data by checking the correlation of the innovation sequence. All of the above are achieved based on the measurements and suitable modelling of the state and measurement equations. The above aspects are discussed in [12, 31].

3.4. Kalman filter and some of its variants

For nonlinear systems, even if the initial distribution is assumed normal, it gets distorted after propagation and so a suitable local approximation or quasi linearization has to be made. In the

EKF the nonlinear systems and/or measurements equations are approximated by appropriate first order Taylor series expansion. The probability density is approximated by a Gaussian, which may distort the true structure and at times could lead to the divergence between the filter prediction and the measurements. In the Unscented Kalman filter (UKF) by Julier et al. [35] approach, instead of linearizing the functions, a set of chosen points are propagated through the nonlinear transformation. These points are so chosen such that the mean, covariance, and possibly also higher order moments match better with the propagated distribution. At an update, only the estimate and covariance matters.

The particle filtering by Gordon et al. [36] is a Monte Carlo technique for state estimation that can handle nonlinear models together with non-Gaussian noise. Here, the state probability density is approximated by using point particles having positive weights. Based on the initial distribution, the weights are chosen and then the particles are propagated following the system dynamics together with the state noise. Then using the measurement, their weights are adjusted and normalised among all the particles. The particles that can track the measurements gain weight and the ones far away lose their weights. However, after a while, all but one weight will become zero leading to degeneracy. A resampling scheme is introduced to solve the degeneracy problem that discard the particles with small weights and focus on the particles with more significant weights. Then, the procedure continues sequentially over the measurements.

For large size systems, such as those occurring in geophysical studies maintaining the covariance matrix computationally being difficult, in the ensemble KF (EnKF) for large problems Evensen [24], the estimate and the covariance matrix are replaced by the sample covariance from a large number of ensemble members similar to a particle in the particle filter. Each member of the ensemble is propagated including the process noise and later updated using a so-called virtual observation. Again, the procedure continues sequentially over the measurements.

One may note the evolution of the variants of the Kalman filter possesses some similarities as it progressed to handle simple, complex, to massive problems as in many other fields such as fluid dynamics or structural mechanics. In these cases, commencing from simple geometries one obtains closed form analytical solutions as in the linear KF, wherein the gains can be pre computed to process the data as and when they arrive. Then for involved nonlinear state and measurements one uses local linearization and numerical calculations as in EKF. When the geometry is complex and the boundary conditions are involved, it becomes necessary to discretise and form cells over appropriate space and time as in particle filtering to obtain the solution. Further when massively complex geometries and boundary conditions occur other innovative approaches like Ensemble Kalman Filter have been developed. An extensive bibliography of the nonlinear estimation is provided by Georgios [37] and an excellent review of nonlinear filters is given by Daum [38].

4. Tuning of the Kalman filter statistics

The solution for the linear filtering problem in discrete time was proposed in the famous 1960 paper by Kalman [4]. This was followed for continuous time in 1961 by Kalman and Bucy [5].

Not many know that the enthusiasm that followed soon after Kalman introduced his filter was damped, the solution was only formal and the statistics of the process **Q** and measurement noise **R** had to be specified to design and implement the filter. After the 1960 paper of Kalman by the time of Gelb's book [39] in 1974 most filter approaches, applications and numerical procedures except the tuning of the filter statistics were in place. Kalman when he proposed the filter dealt with only state estimation. In many present day applications, one does not even know the structure of the state and measurement equations as well as the parameters in them and the statistical characteristics of the state and measurement noise. One can also add the unknown initial state conditions **X0**. All the unknowns have to be estimated using the measurements only. The estimation of the system parameters **Θ**, **X0, P0,** together with **Q** and **R** is called filter design or filter tuning.

An interesting feature of KF is that one can use it to start with at least, without understanding the derivation can tune the filter by trial and error procedures for the statistics without carrying out an optimization as mentioned by Sorenson [40]. Even though the estimator performance may be satisfactory for some '*a priori*' reasonable choice of **P0, Q** and **R**, it could lead to unacceptable results in many cases. Rarely, sensitivity studies on **R** and **Q** like by Subbaraju et al. [41] while estimating the drag in the presence of thrust of Satellite Launch Vehicles are reported. Most reports and publications write out detailed filter equations but the tuning procedures are not spelt out. In fact, the ghost of filter tuning chases without exception every variant or formulation of the Kalman filter. If not tuned properly, it is difficult to infer if the performance of the filter is due to its formulation or filter tuning! It is surprising that most text books on Kalman filtering provide a scanty treatment of the problem of filter tuning that is at the heart of KF design.

One has to tune the statistics **P$_0$, Q** and **R** for a satisfactory filter operation and even now this is generally done manually! Usually, the filter statistics are tuned off line using simulated data and subsequently used for on line and real time applications with some modifications. In spite of its immense applications for more than five decades in many problems of science and technology, the filter tuning has not matured to an easily implementable approach even to handle a constant signal with measurement noise! Generally the tuning is manual or with ad hoc quick fix solutions such as limiting **P** from going to zero, or adding **Q** to increase **P** before calculating the gain and multiplying **P** by a factor to limit **K**, all have obviously limitations in handling involved problems or scenarios. All the above introduce additional parameters to be adjusted that varies for every problem. The present work provides a generalised heuristic approach together with consistency check.

4.1. Qualitative features of P0, Q, and R the filter statistics

Should the **P0 ≡ Q ≡ 0** then the filter will not learn anything from the measurements at all which will be ignored. The **P0** is tricky and generally the off-diagonal elements are set to zero and the diagonal elements are set to large values. However their relative values are crucial for an optimum filter operation. The **R** can be determined from the measured data. In fact if one is satisfied with the measurement accuracy then no filter is needed. The main activity of the filter is to follow the measurements and further reduce or suppress the effect of noise. In spite of being labelled as 'notorious' it is only the **Q** that an analyst can estimate, account or offset for

some deficiency, inaccuracy, or error in the following namely in the initial conditions, the unmodelled or unmodelable errors in the system and measurement equations, control or external input, measurement noise statistics, the errors in the numerical state and covariance propagation, a small increase in $\mathbf{Q}$ to offset modest errors in Taylor series approximations in EKF(!), or in the update operations of the Kalman filter. The $\mathbf{Q}$ helps to inject uncertainty into the state equations to assist the filter to learn from the measurements and also controls the steady state filter response. Too large a value of $\mathbf{Q}$ will lead to a short transient with large steady state uncertainty of the estimates and a small $\mathbf{Q}$ vice versa.

The $\mathbf{Q}$ is helpful to track systems whose dynamical equations are unknown. Some classic examples are the GPS receiver clocks, satellite, trajectory of aircraft, missiles, and re-entry objects. These are handled by using the kinematic relations as state equations among the position, velocity, acceleration and even jerk [42] driven by white Gaussian noise $\mathbf{Q}$ to enable the filter to track these systems. In general one can simulate any real world dynamical systems. Apart from simulating the dynamical system the process noise inhibits the onset of instability of the filter operation. Though $\mathbf{Q}$ is considered notorious it is the life line of the Kalman filter.

4.2. Choice of X0 and P0 for states and parameters

Since some of the states are generally measured either the first or the average of the first few measurements can be taken as the initial value $\mathbf{X0}$ for the state. The initial parameters values can be guessed if experimental or computational results are available. The $\mathbf{P0}$ is one of the important tuning parameters as stressed by very few like Maybeck [43], Candy [44], Gemson [45], Gemson and Ananthasayanam [46], and Sarkar et al. [47]. Generally a guess $\mathbf{P0}$ tends to become very low after some data points. In order to make the filter learn from the subsequent measurements an additional $\mathbf{Q}$ is introduced into the state equations even when there is no model uncertainty. In the present $\mathbf{RRR}$ a proper $\mathbf{P0}$ without any $\mathbf{Q}$ is shown to be possible for the above. The choice of $\mathbf{P0}$ can affect the final covariance from the filter operation, which can be crucial in certain state estimation problems such as impact point estimation and its uncertainty for target tracking. Even in parameter estimation problems the estimates and their uncertainties can be important in the design of control systems.

4.3. Tuning filter statistics with both R and Q

When the data contains the effect of both $\mathbf{R}$ and $\mathbf{Q}$ it becomes notorious for analysis. With no $\mathbf{R}$ and $\mathbf{Q}$ the system dynamics is exact. The process noise input at various times makes the system to wander randomly. When measurements are made on this wandering dynamical state it is blurred. The smoothing filter provides the best possible state estimate at all time points by suppressing the effect of measurement noise. Hence it is best to consider the smoothed state in order to estimate the process noise.

$$\left(\text{State dynamics with } \mathbf{R} \equiv \mathbf{Q} \equiv \mathbf{0}\right) + \text{Cumulative Effect of } \mathbf{Q} = \text{Smoothed State} \quad (21)$$

$$\text{Smoothed State} + \mathbf{R} = \text{Measured Signal or in other words} \quad (22)$$

$$\text{Measured Signal} + \text{Forward and Smoothing filter} = \text{Smoothed State with } \mathbf{R} \equiv \mathbf{0} \quad (23)$$

Shumway and Stoffer [48] did just this using the Estimation Maximisation (EM) approach. The book by McLachlan and Krishan [49] is well known on EM. Shyam et al. [31] introduced another approach to estimate $\mathbf{Q}$ based on the difference between the stochastic and dynamical trajectory called DSDT which provided statistically similar results to EM.

The interesting point is the filter by tracking the drifted dynamical behaviour with $\mathbf{Q}$, it estimates the parameters controlling the original dynamics of the system without the effect of $\mathbf{R}$ and $\mathbf{Q}$. Since $\mathbf{R}$ and $\mathbf{Q}$ occur, respectively, in the measurement and state equations their effects on the filter are negatively correlated as stated by Bohlin [50]. Thus during simultaneous recursive estimation if the statistics for estimating them are not properly chosen then $\mathbf{R}$ is overestimated and $\mathbf{Q}$ is underestimated and vice versa. This is just the reason in Gemson [45] and Gemson and Ananthasayanam [46] one has to update $\mathbf{R}$ and $\mathbf{Q}$ alternately. The filter operating on the data generates prior, posterior, and smoothed state estimates and their covariances thus helping to generate candidate 'statistic' to estimate $\mathbf{R}$ and $\mathbf{Q}$. In EKF if the unknown noise covariances are incorrectly specified biased estimates can arise. Even when the Θ are known, with an inaccurate $\mathbf{R}$ and $\mathbf{Q}$ the filter may give poor estimates, or even diverge [51].

4.4. Adaptive Kalman filtering approaches

There are broadly four approaches for adaptive filtering namely Bayesian, Maximum Likelihood, Covariance Matching and Correlation Techniques (Mehra [52]) apart from other techniques. The present $\mathbf{RRR}$ falls in the category of covariance matching. Why there are so many formulations for solving an optimization problem? The reason is the unknowns do not occur in a simple way in the cost function, and there are many transformed variables with which one tries to solve for the basic unknowns, further the size and the required compatibility conditions among the transformed variables lead to the many difficulties not generally found in the classical optimization problems. Also many attempts have been made using probabilistic methods. However when the dimension, nonlinearity and the range of search space become large these could become computationally prohibitive and could lead to local minimum. One can summarise that either deterministic or probabilistic optimization approaches do not appear to be easy and general for solving the filter tuning problem. A simple recursive filtering approach was tried and fortuitously it did as will be shown subsequently.

Exact filtering solutions are very hard, approximate choice can lead to inappropriate results but heuristic approaches provide the middle path in designing the Kalman filter like the $\mathbf{RRR}$. An adaptive heuristic approach in general updates the $\mathbf{X0, P0, \Theta, R}$ and $\mathbf{Q}$ at a point, over a window, after a pass or after multiple passes by applying some corrections to them based on changes, iterations or sample statistics such that the numerical solution does not diverge but converges to the best possible estimates. Examples of heuristic approach is by Myers and Tapley (MT) in [53] for $\mathbf{R}$ and $\mathbf{Q}$ and in [45, 46] for $\mathbf{P0}$, $\mathbf{Q}$, and $\mathbf{R}$.

5. Cost functions along with RRR for checking Kalman filter tuning

The present $\mathbf{RRR}$ contains no direct optimization of any cost function. We have purely iterated the filter on the measured data; but after each iteration, the unknowns $\mathbf{X0, P0, Q, R}$ and Θ have

been updated. Such an iterative procedure leads to a converged solution for the estimate as well as the CRB that is very close to that obtained using any optimization method that minimises any suitable cost function. The estimates for the parameter, the covariance for uncertainty, and the noise statistics reach very near their final values in about five to ten iterations that can be extended for higher accuracy. Also, we introduce many cost functions as below to provide an indication about the filter performance and consistency.

Fundamentally, the estimation theory (ET) is an optimization problem. Hence a suitable cost function $\mathbf{J}$ is generally chosen. Essentially, there are two elements in ET: (i) defining a cost function and (ii) adopting a suitable algorithm to minimise the cost function. The general log likelihood cost $\mathbf{L}$ for normally distributed error 'e_k' summed over $\mathbf{N}$ time points is given by Gemson [45]

$$\mathbf{L}(\mathbf{\Theta}) = (1/\mathbf{N}) \sum e_k \, \mathbf{A}_k{}^{-1} \, e_k{}^T \; + \; \log \, (\det(\mathbf{A}_k)) \tag{24}$$

where $\mathbf{A_k}$ is the error covariance matrix and $\det(\mathbf{A})$ represents determinant of matrix $\mathbf{A}$. It may be noted the parameters $\mathbf{\Theta}$ occur implicitly in the cost function $\mathbf{L}$. From the e_k of the first term, one can see the principle of weighted least squares, and $\mathbf{L}$ for N-1 and N terms the sequential least squares with the constraint of the state equation as mentioned by Sorenson [40].

Further based on many available statistics from the filter operation it is possible to define many more cost functions (not written here for brevity but available in Shyam et al. [31]), but only describe them. The cost $\mathbf{J0}$ is based on squaring $[(\mathbf{X0}–\mathbf{X}(t0)]$ and scaled with respect to $\mathbf{P0}$ shows how well the initial conditions are balanced. Similarly the set of costs $(\mathbf{J1}, \mathbf{J2}, \mathbf{J3})$ and $(\mathbf{J6}, \mathbf{J7}, \mathbf{J8})$ are respectively derived from the sets $(\mathbf{v_k}, \mathbf{v_f}, \mathbf{v_s})$ and $(\mathbf{w}_1(k|\mathbf{N}), \mathbf{w}_2(k|\mathbf{N}), \mathbf{w}_3(k|\mathbf{N}))$ which are different estimates for local measurement and process noise samples. These indicate how well the measurement and state equations are balanced and should tend to the number of measurement and state equations for a good solution. When $\mathbf{Q} \equiv 0$, the cost $\mathbf{J4}$ is the difference between measurement and state dynamics based on the estimated parameter and is expected to tend towards the trace of $\mathbf{R}$. The cost $\mathbf{J5}$ equals $\mathbf{J1}$ + $\log \, (\mathbf{det}(\text{cov}(\mathbf{v_k}))$ is the negative of the log likelihood based on the innovation. All such costs indicate how well the state and measurement equations are balanced, and further the estimates and the covariances both given by the filter are consistent as well. One can formulate any number of cost functions to estimate the parameters and the filter statistics. But it is not possible to estimate the true value of the unknowns but be only around them due to statistical fluctuations percolating all over the unknowns.

5.1. Probability matching prior interpretation for P0

This interpretation was given earlier in the paper by Ananthasayanam et al. [54], and used by Shyam et al. [55]. As mentioned earlier the importance of $\mathbf{P0}$ has not been much appreciated in the literature on ET and more so in Kalman Filtering though statisticians have been discussing the philosophical and practical differences between Frequentist and Bayesian approach. The deterministic Newton-Raphson (NR) optimization of a cost function approach in Ananthasayanam et al. [56] provides Frequentist results and the Kalman filtering approach is the Bayesian route. Consider the simple case of a constant signal with measurement noise. In the Frequentist approach the calculation of the mean and standard deviation and the noise is

simple. However in the Bayesian approach the above result is not reachable unless a proper **P0** is also chosen. The choice of appropriate **P0** is the probability matching prior (**PMP**) providing a bridge between the above approaches. With a large amount of data the differences in the results from the above approaches vanish. Since **PMP** is not unique its choice depends on the purpose. Presently **P0** is chosen to obtain proper estimates and CRBs for the unknown parameters as well as the noise statistics **R** and **Q**. The success of **RRR** has been due to the choice of **P0** by scaling and further trimming it. Further in addition the simultaneous choice of appropriate statistics for **R** and **Q** has made using the many filter statistics available after every filter pass using the EM approach. When $\mathbf{Q} \equiv 0$ the choice for **R** is easy but when $\mathbf{Q} > 0$ since the Kalman filter is compulsory in both approaches we look for consistency based on simulated studies by comparing the statistical characteristics of the injected and estimated **R** and **Q** noise sample sequences. Further the various cost functions introduced earlier in **RRR** help to obtain confidence in the results and more so while analysing real flight test data. Since the present **RRR** is believed to provide near optimum but not an exact solution it is called as a 'reference' and not a 'standard'.

5.2. Choice of X0 and P0 in RRR

Commencing from an assumed reasonable initial choice for **X0**, **P0**, Θ, **R** and **Q** the first filter pass through the data is made. Then a backward smoothing is carried out using the Rauch et al. [29] smoother. The smoothing leads to the best possible state and parameter estimates and their covariances based on all the data. After smoothing the state estimates and their covariances change but not those of the parameters. If one uses the smoothed initial state covariance $\mathbf{P(0|N)}$ and use it as the **P0** for the next pass then the final covariance will keep on decreasing with further filter passes and eventually tend towards zero. In order to overcome this, the final covariance at the end of the pass was scaled up by Shyam et al. [31] by **N** to provide **P0** at the beginning of the next pass:

$$\mathbf{P0} = \mathbf{N} \times \mathbf{P(N|N)} \tag{25}$$

A heuristic reasoning from statistics is that the mean from a sample has an uncertainty **P** that keeps decreasing with sample size as **P/N** where **P** is the population variance. Since, in the filter steps, the estimates and their update refer to the sample and the other covariance propagation and their update and the calculation of the Kalman gain refers to the ensemble characteristics before every filter pass, we carry out the above scale up to obtain the **P0** (after further trimming as well) for the next filter pass.

5.3. Estimation of R and Q using the EM/DSDT, MS and MT methods

We now provide some estimates for the measurement noise covariance **R**. Bavdekar et al. [57] use the extended EM given by:

$$\mathbf{R} = (1/\mathbf{N}) \sum_1^N \boldsymbol{\nu}_k \, \boldsymbol{\nu}_k^T + \mathbf{H}(k|\mathbf{N})\mathbf{P}(k|\mathbf{N})\mathbf{H}(k|\mathbf{N})^T \tag{26}$$

The choice of Mohamed and Schwarz (MS) in [58] based on filter residue is

$$\mathbf{R} = (1/\mathbf{N}) \sum_1^N \boldsymbol{\nu}_f \, \boldsymbol{\nu}_f^T + \mathbf{H}(\mathrm{k}|\mathrm{k})\mathbf{P}(\mathrm{k}|\mathrm{k})\boldsymbol{H}(k|k)^T \tag{27}$$

The choice of Myers and Tapley (MT) in [50] based on innovation is

$$\mathbf{R} = (1/\mathbf{N}) \sum_1^N \boldsymbol{\nu}_s \, \boldsymbol{\nu}_s^T - \mathbf{H}(\mathrm{k}|\mathrm{k}-1)\mathbf{P}(\mathrm{k}|\mathrm{k}-1)\boldsymbol{H}(k|k-1)^T \tag{28}$$

Bavdekar et al. [57] use the smoothed statistic $w_1(\mathrm{k}|\mathbf{N}) = \mathbf{X}(\mathrm{k}|\mathbf{N})\mathbf{-f}(\mathbf{X}(\mathrm{k}-1|\mathbf{N})\,)$ in EM estimate as

$$\begin{aligned}
\mathbf{Q} = (1/\mathbf{N}) \sum_1^N \Big\{ & \boldsymbol{w}_1(k|\mathbf{N}) \, \boldsymbol{w}_1(k|\mathbf{N})^T + \mathbf{P}(k|\mathbf{N}) + \mathbf{F}(\mathrm{k}-1|\mathbf{N})\mathbf{P}(k-1|\mathbf{N}) \, \boldsymbol{F}(k-1|\mathbf{N})^T \\
& -\mathbf{P}(k,k-1|N)\boldsymbol{F}(k-1|\boldsymbol{N})^T - \boldsymbol{P}(k,k-1|\boldsymbol{N})^T\mathbf{F}(k-1|\mathbf{N}) \Big\}
\end{aligned} \tag{29}$$

The DSDT statistic for $\mathbf{Q}$ is given in Shyam et al. [31]

$$\begin{aligned}
\mathbf{Q} = \left(\frac{1}{N}\right) \sum_1^N \Big\{ & w_2(k|\boldsymbol{N}) \, w_2(k|\boldsymbol{N})^T + \mathbf{P}(k|\boldsymbol{N}) + \mathbf{Fd}(k-1|\boldsymbol{N})\mathbf{P}(k-1|\boldsymbol{N})\boldsymbol{Fd}(k-1|\boldsymbol{N})^T \\
& -\mathbf{P}(k,k-1|\boldsymbol{N})\boldsymbol{Fd}(k-1|\boldsymbol{N})^T - \boldsymbol{P}(k,k-1|\boldsymbol{N})^T\mathbf{Fd}(k-1|\mathbf{N}) \Big\}
\end{aligned} \tag{30}$$

where $\mathbf{w}_2(\mathrm{k}|\mathbf{N}) = \mathbf{X}(\mathrm{k}|\mathbf{N})\mathbf{-Xd}(\mathrm{k}|\mathbf{N})\mathbf{-Fd}(k-1|\mathbf{N})(\mathbf{X}(k-1|\mathbf{N})\mathbf{-Xd}(k-1|\mathbf{N}))$ and with $\mathbf{Xd}(0|\mathbf{N}) = \mathbf{X}(0|\mathbf{N})$ is the predicted state trajectory without measurement and process noise using the estimated parameter $\boldsymbol{\Theta}(\mathbf{N}|\mathbf{N})$. The $\mathbf{P}(\mathrm{k}|\mathbf{N})$ is the smoothed covariance and $\mathbf{P}(\mathrm{k}, \mathrm{k}-1|\mathbf{N})$ is the lag-one covariance for k = **N-1**, **N-2**, …1.

The Mohamed and Schwarz [58] estimated $\mathbf{Q}$ in terms of the innovation and the smoothed gain $\mathbf{K}(\mathrm{k}|\mathbf{N})$ based on [29] by

$$\mathbf{Q} = \mathbf{K}(\mathbf{k}|\mathbf{N}) \left\{ \sum_1^N \boldsymbol{\nu}_s \, \boldsymbol{\nu}_s{}^T \right\} \mathbf{K}(k|N)^T \tag{31}$$

The choice of Myers and Tapley [53] for $\mathbf{Q}$ with $\mathbf{w}_3(\mathrm{k}|\mathbf{N}) = \mathbf{X}(\mathrm{k}|\mathrm{k}) - \mathbf{X}(\mathrm{k}|\mathrm{k\text{-}1})$ is

$$\mathbf{Q} = \left(\frac{1}{N}\right) \sum_1^N \left\{ w_3(\mathrm{k}|\mathrm{k}) \, w_3(\mathrm{k}|\mathrm{k})^T - \left(\mathbf{F}(k-1|\mathrm{k}-1)\mathbf{P}(\mathrm{k}-1|\mathrm{k}-1)\mathbf{F}(k-1|\mathrm{k}-1)^{\mathrm{T}} - \mathbf{P}(\mathrm{k}|\mathrm{k}) \right) \right\} \tag{32}$$

All the process noise samples, $\mathbf{w}_1(\mathrm{k}|\mathrm{k})$, $\mathbf{w}_2(\mathrm{k}|\mathrm{k})$, and $\mathbf{w}_3(\mathrm{k}|\mathrm{k})$, are assumed to be of zero mean. It turns out that the smoothed statistics $\mathbf{w}_1(\mathrm{k}|\mathrm{k})$ and $\mathbf{w}_2(\mathrm{k}|\mathrm{k})$ based on EM and DSDT, respectively, are very close and either can be used for $\mathbf{Q}$ estimation.

5.4. The RRR method for the EKF

The following steps explain the recursive or iterative **RRR** algorithm for the EKF:

1. Given the system model and the measurements, the first filter pass through the data of EKF is carried out using guess values of **X0**, **P0**, **Θ**, **R** and **Q**.

2. The RTS smoother is used backwards to get smoothed state and covariance estimates.

3. If **X0** is unknown, then the smoothed state values can be used as the initial state values.

4. The estimated smoothed **P0** is scaled up by the number of time points **N** and further all elements except the diagonal terms corresponding to the parameters are set to zero. Due to the effect of statistical percolation effect, the estimated **R** and **Q** will in general be full. But, only the diagonal terms in **Q** need to be used in the basic state equations and not in the parameter states. Only the diagonal terms in **R** need to be used in the measurement equations. These are summarised as below. The quadrant on the upper left denotes the state, the bottom right the parameter states, and the others the cross terms. The (0) below shows the null and D the diagonal matrices. This is followed for all iterations.

$$\mathbf{P0} = \begin{bmatrix} 0 & 0 \\ 0 & D \end{bmatrix}; \ \mathbf{Q} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}; \ \mathbf{R} = [D] \tag{33}$$

5. Then, the filter is run again using the above updates of **X0**, **P0**, **Θ**, **Q** and **R** till statistical equilibrium is reached.

6. Different cost functions (**J1–J8**) are checked for convergence.

The convergences of the following quantities after all the iterations are analysed:

1. The parameter estimates **Θ** and their covariances **P(Θ)**.

2. The sample noise sequences $\mathbf{v_k}$, $\mathbf{v_f}$, $\mathbf{v_s}$ and of $\mathbf{w_1}$, $\mathbf{w_2}$, and $\mathbf{w_3}$ with ±1 sigma bounds, their autocorrelations and **Q** and **R**.

3. The state dynamics **Xd** (with $\mathbf{R} \equiv \mathbf{Q} \equiv 0$) based on the estimated **Θ**, smoothed **X(k|N)** and the measurement **Z**(k).

4. The various cost functions **J1–J8** after the final convergence.

5.5. Some remarks on running the RRR

If the value of **Q** for any state is known to be zero, then the value of **Q** is set at 10^{-10} or lower for all iterations to help the filter, which would otherwise generate a pseudo-**Q**, and then slowly grind it to zero in thousands of iterations. For $\mathbf{Q} \equiv 0$ case, one can estimate **R** even by ignoring the second order covariance terms. It is of interest to note that for $\mathbf{Q} > 0$ case unless the second order covariance terms are also included in the estimate for both **R** and **Q** the **RRR** does not converge.

6. Simulation study of a spring, mass, and damper (SMD) system

The **RRR** is first applied to a very simple spring, mass, and damper system with $\mathbf{R} \equiv \mathbf{Q} \equiv 0$. For such a situation, the Newton Raphson optimization of the innovation cost [56] served as a bench mark for tuning the filter statistics in **RRR** (with no cost optimization!) to get the closest possible estimates and the CRB. Later, when **Q** is included, we looked for the consistency between the injected **R** and **Q** noise sequences and their statistics. The SMD system with weak nonlinear spring constant in time (t) is given by

$$\dot{x}_1(t) = x_2(t); \; \dot{x}_2(t) = - \Theta_1 x_1(t) - \Theta_2 x_2(t) - \Theta_3 x_1^3(t) \tag{34}$$

where x_1 and x_2 are the displacement and velocity state with initial conditions 1 and 0, respectively. The 'dot' represents differentiation with respect to time (t). The unknown parameter vector is $\Theta = [\Theta_1, \Theta_2, \Theta_3]^{\mathbf{T}}$ with true values $\theta_{\mathbf{true}} = [4.0, 0.4, 0.6]^{\mathbf{T}}$. Θ_3 is a weak parameter and does not affect the system dynamics much. The complete state vector $\mathbf{X} = [x_1, x_2, \Theta_1, \Theta_2, \Theta_3]^T$ is of size (5 x 1). The measurement equation is given by

$$\mathbf{Z}(k) = \mathbf{H}\,\mathbf{X}(k) \tag{35}$$

where $\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ is the measurement matrix. The noise covariances are **R** = diag (0.001, 0.004) and **Q** = diag (0.001, 0.002), thus keeping the noise levels of same order to test the robustness of the **RRR**. The initial guess value of **P0**, **Q**, were chosen, respectively, as diagonal with all values as 10^{-1}, and $\mathbf{R} = 2^{-1}$ for all measurement channels. The initial Θ was chosen within $\pm 20\%$ error of the true values. A total of **N** = 100 measurement data are simulated with the time varying from 0 to 10 s in very small steps of dt = 0.1 s. For zero process noise case, the maximum number of iterations is set to 20 over 50 simulations, and for nonzero process noise case, it is set to 100 over 50 simulations for obtaining generally four digits accuracy (though not necessary). The brief results as presented in **Figures 2** and **3**. In the present **RRR**, it was noticed that generally even if the initial **P0**, **Q**, and **R** were varied over a wide range of powers from -3 to $+3$ (or even more) together with all the initial Θ parameter set to zero leads to the same estimation results for a given data showing its robustness. Further, it may be noted that for all simulated and real data analysed and reported in [31] even if all the initial state and parameters were set to zero the solutions converged to the appropriate values!

6.1. Analysis of simulated SMD data

The **RRR**, when compared to [53, 57, 58] has hardly any instability when processing the simulated data, reaching statistical equilibrium in around 20 iterations. This can be seen from **Figure 2** for Θ, their CRBs (= **P0/N**), **R**, **Q**, **J1–J8.** The cost functions **J1–J8** provides confidence in the results and to compare **RRR** with other approaches. **Figure 3** shows the measurements wrap around the smoothed estimates. More details are available in Shyam et al. [31]. For further insight, the filter results with 500 iterations shown in **Figures 4** and **5** indicate three phases. Firstly, from a
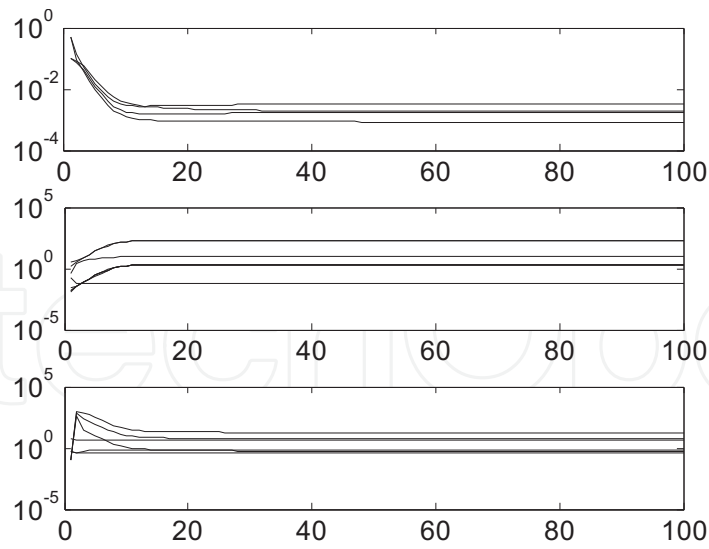
Figure 2. (i) Variation of **Q** and **R**, (ii) initial parameters **Θ**, and **P0**, and (iii) costs **J1–J8** with iterations.

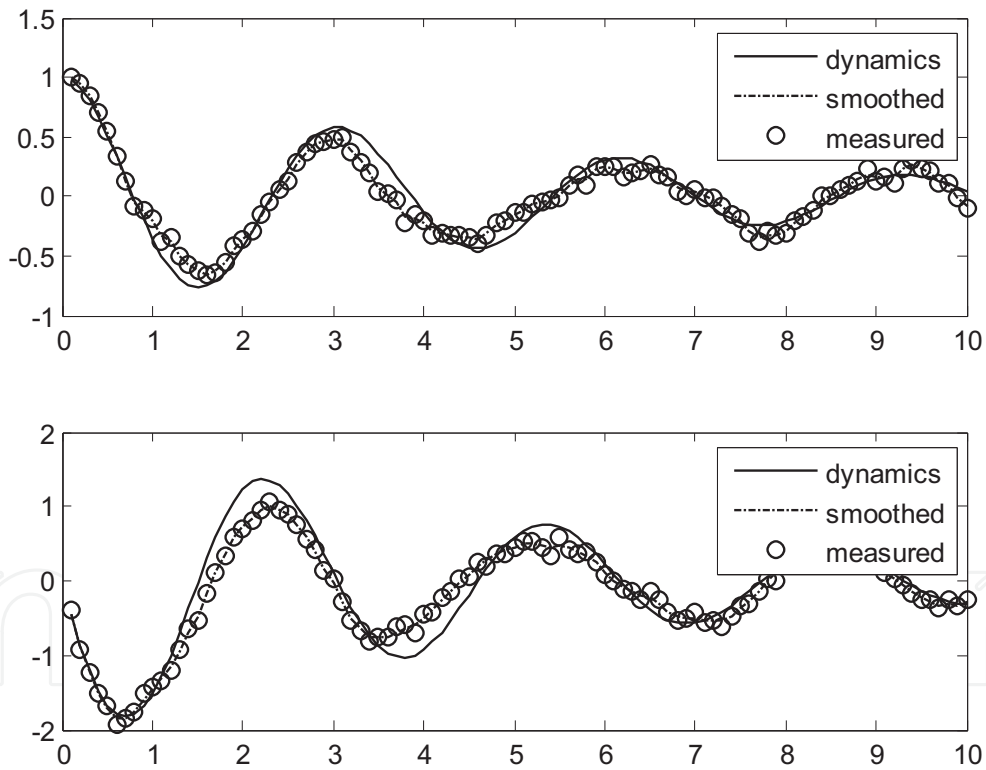**Figure 3.** **Xd**, **X**(k|N), and **Z**(k) by (o) for displacement and velocity versus time (s).

nonequilibrium state the filter statistics reach statistical equilibrium in 10–20 iterations enough for use in practice. The next phase shows the second moments **R** and **Q** converge earliest followed by Θ and finally their CRBs. This is at variance with the fact mentioned in Section 3.2, namely lower order moments converge faster than higher order moments for samples from a distribution. But the statistics derived here from the Kalman filter are not simple. Finally, there is equilibrium with numerical fluctuations depending on computer accuracy.
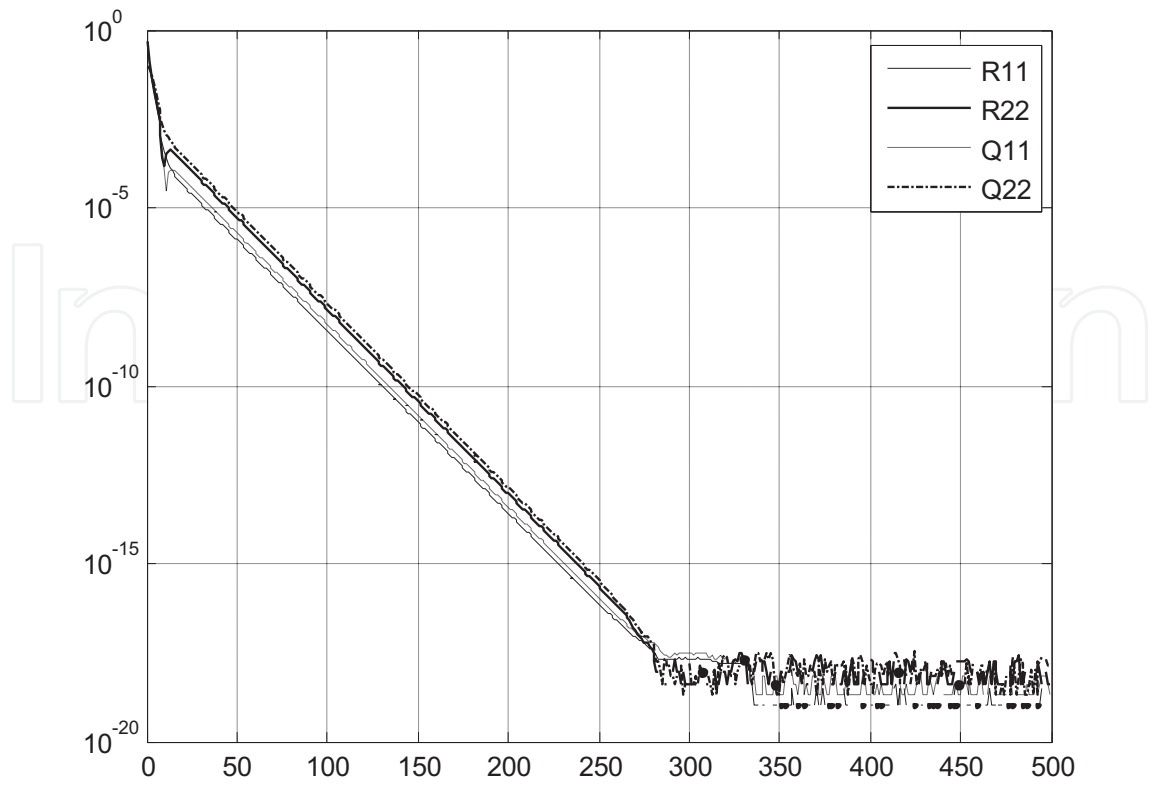
Figure 4. Difference of the diagonal elements of **R** and **Q** from their values after 500 iterations.
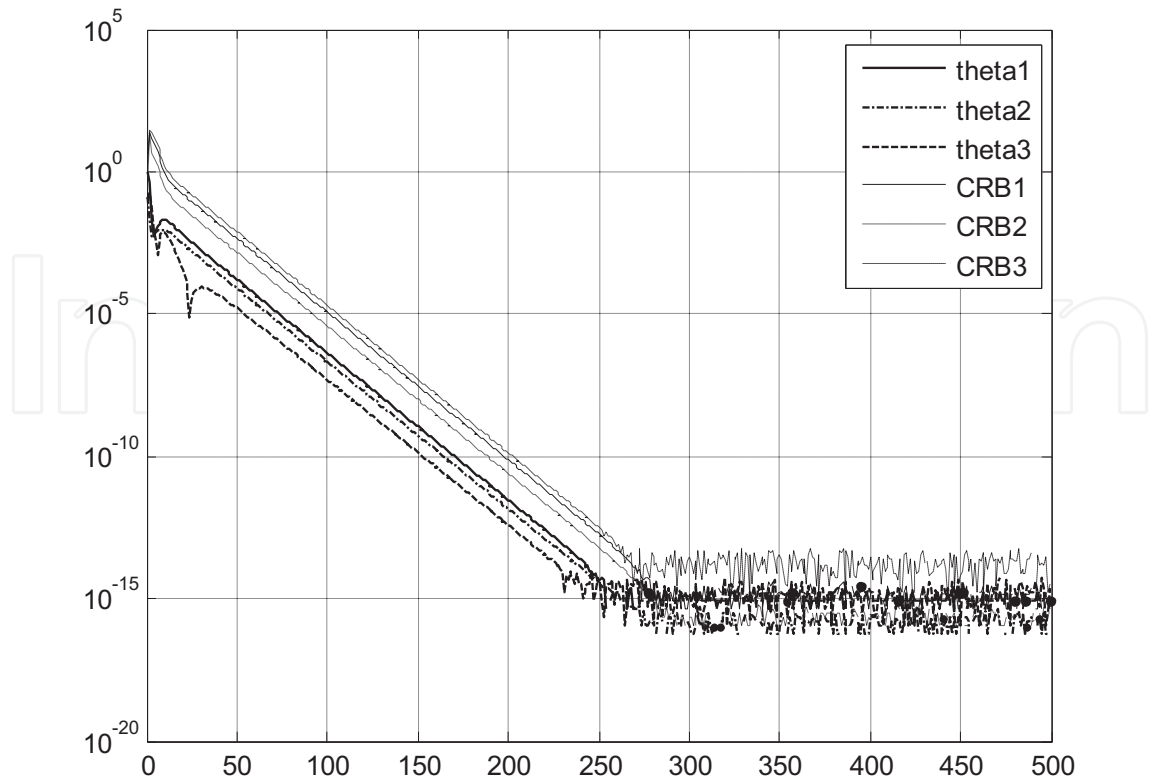

Figure 5. Difference of parameters **Θ** and their CRBs from their values after 500 iterations.

6.2. Analysis of real airplane flight test data

The real airplane flight tests cannot always be conducted in an ideal situation of **Q** and **R** being white Gaussian. The measurements may not be at the center of gravity, possess bias and scale factors, which have to be modelled and estimated. The coupling between the longitudinal and lateral motion brings in difficulty but makes the problem interesting. At times, the noisy measurements from the lateral motion are fed into the longitudinal states, and thus are input as process noise. This is an example of introducing subjectivity in estimation theory. The real data set is obtained along with notations from [59]. There is a peculiar manoeuvre, where the elevator angle (δ_e in deg) shown in **Figure 6** is imparted when the aircraft (T 37B) is rolling through a full rotation during aileron roll. The coupling between the longitudinal and lateral motion is replaced by their measured values, namely the roll angle (φ_m), sideslip (β_m), velocity (V_m), roll rate (p_m), yaw rate (r_m) and the angle of attack (α_m) as shown in **Figures 6–8**. The state equations (n = 3) for the angle of attack (α), pitch rate (q), and the pitch angle (θ), respectively, are

$$
\dot{\alpha} = -\frac{\overline{q}\,S}{mV_m\,Cos\,(\beta_m)}\left(C_{L_\alpha}\,\alpha + C_{L_{\delta_e}}\,\delta_e + C_{L_0}\,\right) + q + \frac{g}{V_m\,Cos\,(\beta_m)}\left(\cos\,(\phi_m)\left(\cos\,(\alpha_m)\cos\,(\theta)\right.\right.
$$
$$
+ \sin\,(\alpha_m)\,\sin\,(\theta)\right) - \tan\,(\beta_m)\left(p_m\cos\,(\alpha_m) + r_m\,\,\sin\alpha_m\right))
$$

(36)

$$
\dot{q} = \frac{\overline{q}S\overline{c}}{I_{yy}}\left(C_{m_\alpha}\alpha + C_{m_q}\frac{\overline{c}}{2V}q + \,C_{m\dot{\alpha}}\,\frac{\overline{c}}{2V}\,\dot{\alpha} + C_{m_{\delta_e}}\,\delta_e + C_{m0}\,\right) + \frac{I_{zz} - I_{xx}}{I_{yy}}\,\,r_m\,p_m
$$

(37)

$$
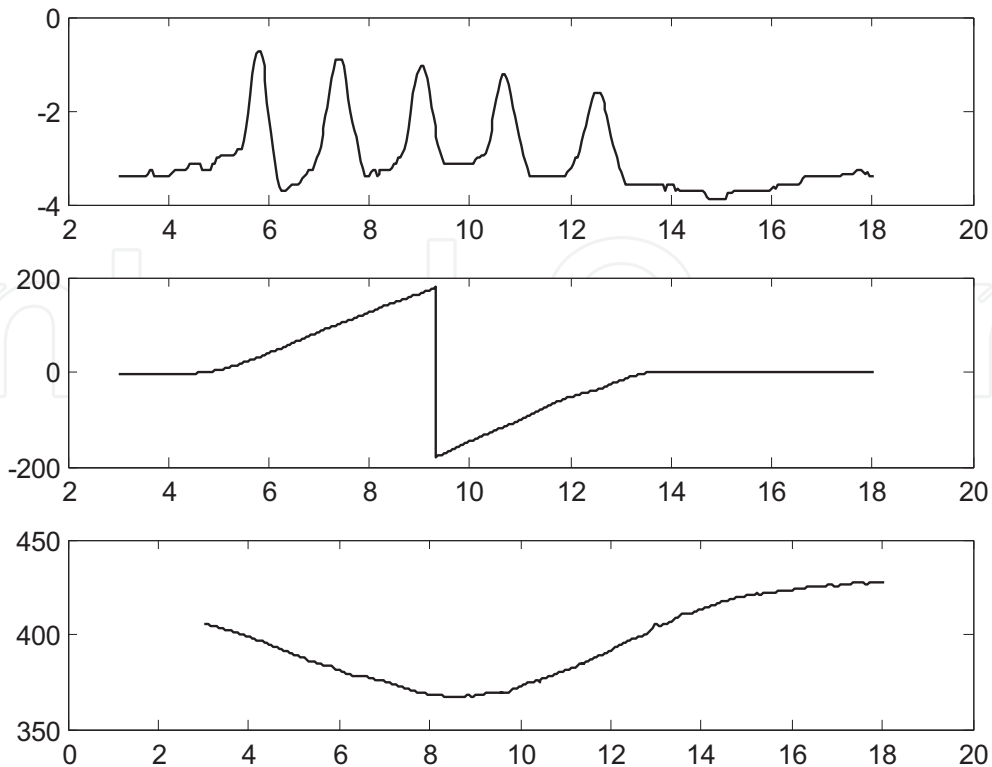\dot{\theta} = q\cos\,(\phi_m) - r_m\sin\,(\phi_m) + \theta_0
$$

(38)

Figure 6. Control input δ_e in deg, roll angle ϕ_m in deg, velocity V_m in ft/sec versus time (s).
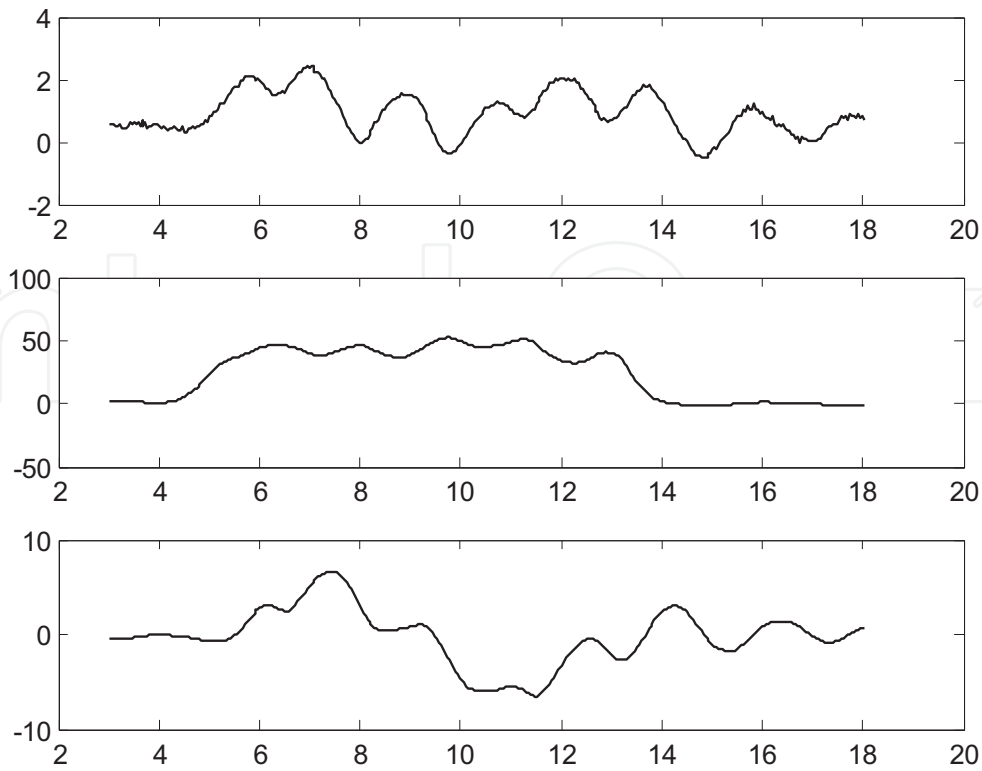
Figure 7. Sideslip angle β_m in deg, roll rate p_m in deg/sec, and yaw rate r_m in deg/sec, versus time (s).
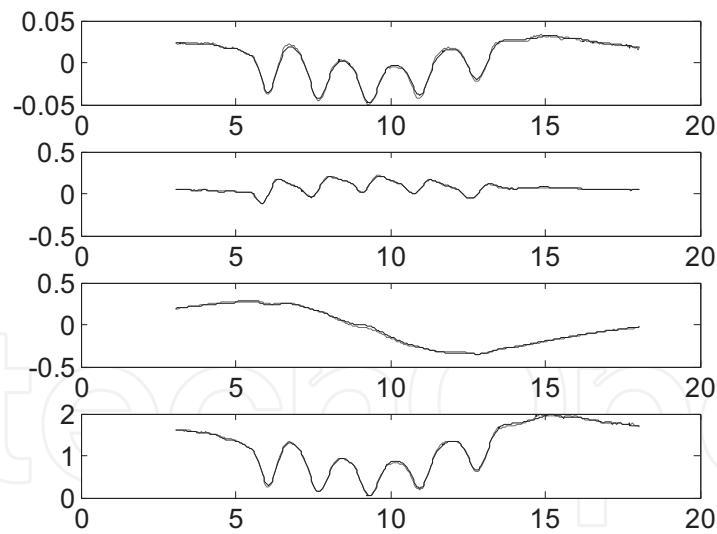

Figure 8. **Xd**(cont), and **Z**(dashdot) for α (deg), Q (deg/sec), θ (deg), and a_n with time (s).

The measurement equations (m = 4) for angle of attack, pitch rate, pitch, and normal accelera-
tion are given by

$$\alpha_m = K_\alpha \, \alpha - K_\alpha x_\alpha \frac{q}{V}; \quad q_m = q \; ; \theta_m = \theta \tag{39}$$

$$a_{n_m} = \frac{\overline{q}S}{mg} \left(C_{N_\alpha}\alpha + C_{N_{\delta_e}}\delta_e + C_{N_0} \right) + \frac{x_{a_n}}{g} \, \dot{q} \tag{40}$$

The unknown parameters (p = 10) are $\left(C_{L_\alpha}, C_{L_{\delta_e}}, C_{L_0}, C_{m_\alpha}, C_{m_q}, C_{m\dot\alpha}, C_{m_{\delta_e}}, C_{m_0}, \theta_0, C_{N_0}\right)^{\mathrm{T}}$ with the approximation $C_{N_\alpha} = C_{L_\alpha}$ and $C_{N_{\delta_e}} = C_{L_{\delta_e}}$. The suffix δ_e denotes control derivatives, and suffix zero refers to biases and all others are aerodynamic derivatives. The initial states are taken as the initial measurements and the initial parameter values are taken as (4, 0.15, 0.2, −0.5, −11.5, −5, −1.38, −0.06, −0.01, 0.2)$^{\mathrm{T}}$. The other constant values used are S = 184, m = 196, I_{xx} = 6892.7, I_{yy} = 3952.3, I_{zz} = 10416.4, g = 32.2, $\bar{c}$ = 5.58, $K_\alpha x_\alpha$ = −0.0279, x_{a_n} = 0.101, and K_α = 1.

6.2.1. Remarks on the real data case results

All the real data studies were run for 100 iterations using the **RRR** with **Q** > 0, since the off-diagonal elements of the correlation coefficient matrix C for parameter estimates reduced substantially than for **Q** ≡ 0. **Figures 6–9** show the various input and output quantities from the **RRR**, and **Table 2** provides a comparison of the parameter estimates along with their CRBs
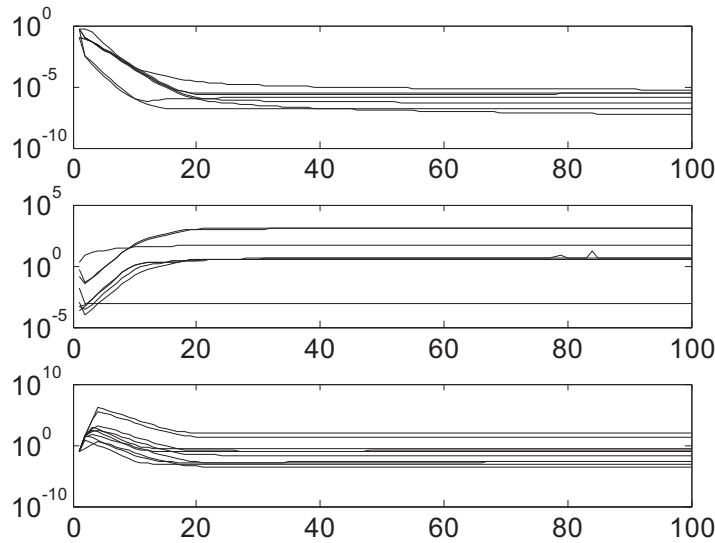

Figure 9. (i) **Q** and **R**, (ii) initial **Θ**, and **P0**, and (iii) the costs **J1–J8** with iterations.

Θ	RRR	NASA(Q≡0)	Gemson	MT	MS
C_{N_0}	0.2538 (0.0014)	0.2541 (0.008935)	0.2503 (0.0014)	0.2540 (0.0016)	0.2635 (0.0026)
C_{L_0}	0.2409 (0.0021)	0.2448 (0.009215)	0.2529 (0.0018)	0.2408 (0.0023)	0.2517 (0.0027)
C_{L_α}	4.9235 (0.0164)	5.1068 (0.1322)	4.9028 (0.0168)	4.9260 (0.0184)	5.0620 (0.0323)
$C_{L_{\delta_e}}$	0.1554 (0.0271)	0.1909 (0.1602)	0.0879 (0.0267)	0.1587 (0.0302)	0.3594 (0.0508)
C_{m_0}	−0.0425 (0.0009)	−.0505 (0.002655)	−0.0507 (0.0024)	−0.0424 (0.0009)	−0.0447 (0.0006)
C_{m_α}	−0.5293 (0.0079)	−0.6474 (0.02339)	−0.6174 (0.0211)	−0.5285 (0.0082)	−0.5590 (0.0055)
C_{m_q}	−11.8596 (0.2402)	−14.2600 (0.6528)	−18.8339 (0.8379)	−11.8255 (.2483)	−12.5965 (0.1400)
$C_{m\dot\alpha}$	−6.8959 (0.4891)	−8.2700 (1.296)	−7.1290 (1.544)	−6.8798 (0.5062)	−6.6713 (0.3021)
$C_{m_{\delta_e}}$	−0.9731 (0.0177)	−1.1614 (0.05371)	−1.1841 (0.471)	−0.9711 (0.0184)	−1.0247 (0.0129)
θ_0	0.0003 (0.0021)	−0.01177 (0.02528)	−0.0037 (0.001)	0.0002 (0.0011)	−0.0006 (0.0007)

Table 2. Real flight test data results (**Θ, CRB(Θ)**).

Method	Measurement noise covariance R $\times 10^{-6}$				Process noise covariance Q $\times 10^{-6}$		
RRR	1.241	0.051	0.460	5.668	0.180	2.954	2.646
MT	1.614	0.240	2.316	2.929	0.203	3.153	0.667
MS	3.160	37.242	9.341	841.55	0.00005	0.0003	0.2386
Cost functions **J1–J8**							
RRR	3.934	4.223	3.616	0.0008	−44.1347	2.975	2.976	2.907
MT	3.766	4.519	3.838	0.0008	−43.7340	4.2266	4.228	2.949
MS	3.162	3.151	2.590	0.0007	−38.0517	8.477	8.466	3.022

Table 3. Real flight test data results (**R, Q, J**).

(in parenthesis) for the real data from different approaches. Since MT and MS do not specify **P0**, the one in **RRR** is used. It turns out that even the parameters $C_{L_{\delta_e}}$ and C_{m_q} strongly affecting the airplane dynamics is estimated vary widely among the various approaches. Also, even if Θ are good the CRBs differ among the approaches such as for C_{N_0} and $C_{m\dot\alpha}$.

We noted earlier about the negatively correlated behaviour of **R** and **Q**. From **Table 3**, it may be noted that the **RRR** relative to other methods generally provides a larger **Q** and a smaller **R.** This implies that the larger Q assists in a better tracking of the wandering state. The smaller **R** tightly wraps around the smoothed state and thus providing generally smaller CRBs. **Figure 8** shows the close match of the measurements with the dynamics in this real case is due to the small **R** and **Q** than in the SMD case. The convergence remarks for Θ, their CRBs (= **P0/N**), **R, Q, J1–J8** as in SMD is valid here as well. The cost functions from **RRR** are generally closer to the number of states and measurements as expected.

The rounded 100(Correlation coefficient) matrix $\mathbf{C_{ij}} = \mathbf{P_{ij}}/sqrt(\mathbf{P_{ii}P_{jj}})$ with **P** denoting the covariance of parameter estimates is.

Θ	C_{N_0}	C_{L_0}	C_{L_α}	$C_{L_{\delta_e}}$	C_{m_0}	C_{m_α}	C_{m_q}	$C_{m\dot\alpha}$	$C_{m_{\delta_e}}$	θ_0
C_{N_0}	100	65	62	98	−13	−10	1	2	−12	0
C_{L_0}	65	100	41	64	−7	−4	−2	5	−5	0
C_{L_α}	62	41	100	67	−8	−19	1	1	−8	0
$C_{L_{\delta_e}}$	98	64	67	100	−13	−11	2	1	−12	0
C_{m_0}	−13	−7	−8	−13	100	88	9	84	99	0
C_{m_α}	−10	−4	−19	−11	88	100	25	70	91	0
C_{m_q}	1	−2	1	2	9	25	100	−27	21	1
$C_{m\dot\alpha}$	2	5	1	1	84	70	−27	100	80	−1
$C_{m_{\delta_e}}$	−12	−5	−8	−12	99	91	21	80	100	0
θ_0	0	0	0	0	0	0	1	−1	0	100

Based on the C from **RRR**, the weakest parameter can be inferred as θ_0, which is uncorrelated with all other parameters and its estimates and uncertainty can vary widely among the approaches. Next, it is possible to group the parameters as (C_{N_0}, C_{L_0}, C_{L_α}, $C_{L_{\delta_e}}$) and (C_{m_0}, C_{m_α}, C_{m_q}, $C_{m\dot\alpha}$, $C_{m_{\delta_e}}$). If a certain state is excited relatively more than others, then the estimated parameter that multiplies it will have lower correlation with other parameters in the set.

The downloadable MATLAB program available in Shyam et al. [31], which has many simulated and real flight test data analysis can be used all the way from teaching, learning to carry out research using Kalman filter.

7. Conclusions

A new approach called **RRR** has been proposed to handle the important problem of tuning the Kalman filter statistics. The importance of **P0** in EKF as the **PMP** is stressed. This along with suitably chosen **R** and **Q** after every filter pass through the data based on various filter statistics converges after few iterations without any optimization. Further, the many cost functions indicate the balance of the state and measurement equations and the consistency of the various filter statistics. These help the user to assess the results and compare **RRR** with other approaches. The efficacy of the **RRR** is demonstrated by application to a simple SMD system and a real airplane flight test data with a large number of unknown aerodynamic parameters.

Acknowledgements

Author details

Mudambi R Ananthasayanam

Address all correspondence to: sayanam2005@yahoo.co.in

Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India

References

[1]   Billard R. Aryabhata and Indian astronomy. Indian Journal of History of Science. 1977;**12**(2): 207-224. Also R. Billard, "L'astronomie Indienne," Publications De Ecole Francaise D'- Extreme-Orient, Paris, 1971

[2]   Sarma KV. Tradition of Aryabhatiya in Kerala: Revision of planetary parameters. Indian Journal of History of Science. 1977;**12**(2):194-199

[3]   Gauss KF. Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium (Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections [Translation]). New York: Dover; 1809. p. 1963

[4]   Kalman RE. A new approach to linear filtering and prediction problems. Transactions of the ASME -Journal of Basic Engineering. 1960;**82**(Series D):35-45

[5]   Kalman RE, Bucy RS. New results in linear filtering and prediction theory. Journal of Basic Engineering. 1961;**83**(1):95-108

[6]   Lauritzen S. Thiele Pioneer in Statistics. Oxford University Press; 2002

[7]   Swerling P. A Proposed Stagewise Differential Correction Procedure for Satellite Tracking and Prediction. Tech. Rep. P-1292. Rand Corporation; 1958

[8]   Stratonovich RL. Application of the theory of Markov process in optimal signal discrimination. Radio Engineering and Electronic Physics. 1960;**1**:1-19

[9]   Plackett RL. Some theorems in least squares. Biometrika. 1950;**37**:149157

[10] Sprott DA. Gauss's contributions to statistics. Historia Mathematica. 1978;**5**:183-203

[11] Wiener N. The Extrapolation, Interpolation and Smoothing of Stationary Time Series: With Engineering Applications. Cambridge, Massachusetts: MIT Press:; New York: Wiley and Sons; London: Chapman & Hall; 1949

[12] Ananthasayanam MR. A relook at the concepts and competence of the Kalman Filter. In: Proceedings of the 42nd AIAA Aerospace Sciences Meeting and Exhibit; 5–8 January 2004; Reno, Nevada. AIAA 2004-571

[13] Klein V, Morelli EA. Aircraft System Identification: Theory and Practice. AIAA Edn. Series; 2006

[14] Jategaonkar RV. Flight Vehicle System Identification: A Time Domain Methodology. Vol. 216. AIAA; 2006

[15] Bar-Shalom Y, Rong Li X, Kirubarajan T. Estimation with Applications to Tracking and Navigation, Theory, Algorithm and Software. John Wiley and Sons; 2000

[16] Ananthasayanam MR, Anilkumar AK, Subba Rao PV. New approach for the evolution and expansion of space debris scenario. Journal of Spacecraft and Rockets. 2006;**43**(6): 1271-1282

[17] Grewal MS, Lawrence RW, Andrews AP. Global Positioning Systems, Inertial Navigation, and Integration. 2nd ed. John Wiley & Sons Inc; 2007

[18] Kleusberg A, Teunissen PJG. GPS for Geodesy. 1st ed. Springer; 1996

[19] Fruhwirth R, Regier M, Bock RK, Grote H, Notz D. Data analysis techniques for high-energy physics. In: Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology; 2000

[20] Federer WT, Murthy BR. Kalman Filter Bibliography: Agriculture, Biology, and Medicine. Technical Report BU-1436-M. Department of Biometrics, Cornell University; 1998

[21] Visser H, Molenaar J. Kalman filter analysis in dendroclimatology. Biometrics. 1988;**44**: 929-940

[22] Wells C. The Kalman Filter in Finance. Springer-Science. Business Media, BV; 1996

[23] Costagli M, Kuruoglu EE. Image separation using particle filters. Digital Signal Processing. 2007;**17**:935-946

[24] Evensen G. Data Assimilation: The Ensemble Kalman Filter. Springer Verlag; 2009

[25] Rao CR. Statistics and Truth: Putting Chance to Work - Ramanujan Memorial Lectures. New Delhi, India: Council of Scientific and Industrial Research; 1987

[26] Samaniego FJ. A Comparison of the Bayesian and Frequentist Approaches to Estimation. Springer Science; Business Media, LLC; 2011

[27] Feller W. An Introduction to Probability Theory and its Applications. 3rd ed. Vol. 1. John Wiley and Sons; 1967

[28] Brown R, Hwang P. Introduction to Random Signals and Applied Kalman Filtering, with MATLAB Exercises. 4th ed. John Wiley and Sons; 2012

[29] Rauch HE, Tung F, Striebel CT. Maximum likelihood estimates of linear dynamic systems. AIAA Journal. 1965;**3**(8):1445-1450

[30] Kailath T. An innovations approach to least-squares estimation part I: Linear filtering in additive white noise. IEEE Transactions on Automatic Control. 2016;**AC-13**(6):646-655

[31] Shyam MM, Naik N, RMO G, Ananthasayanam MR. Introduction to the Kalman Filter and Tuning its Statistics for near Optimal Estimates and Cramer Rao Bound. TR/EE2015/ 401. Kanpur: IIT; 2015 http://arxiv.org/abs/1503.04313

[32] Philip NK, Ananthasayanam MR. Relative position and attitude estimation and control schemes for the final phase of an autonomous docking mission. Acta Astronautica. 2003; **52**(7):511-522

[33] Minsky M. The Society of the Mind. Picador Edition; 1988

[34] Narasimha R. Performance reliability of high maintenance systems. Journal of the Franklin Institute. 1975;**303**:15-29

[35] Julier SJ, Uhlmann JK, Durrant-Whyte HF. A new approach for filtering nonlinear systems. In: Proceedings of the 1995 American Control Conference. Vol. 1995. IEEE Press;

[36] Gordon NJ, Salmond DJ, Smith AFM. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings. 1993;**140**(2):107-113

[37] Georgios BG, Erchin S. bibliography on nonlinear system identification. Signal Processing. 2001;**8**:533-580

[38] Daum F. Nonlinear filters: Beyond the Kalman filter. IEEE A&E Systems Magazine. 2005; **20**:8, Part 2: Tutorials-Daum

[39] Gelb A, editor. Applied Optimal Estimation. Cambridge: Massachusetts: MIT Press; 1974

[40] Sorenson HW. Least squares estimation from gauss to Kalman. IEEE Spectrum. 1970;**7**: 63-68

[41] Subbaraju PV, Ananthasayanam MR, Deshpande SM. Estimation of the Drag of a Satellite Launch Vehicle from Flight Data Using Extended Kalman Filter. Report 88 FM 3, Dept. of Aerospace Engineering. Indian Institute of Science, Bangalore; 1988. Accepted for presentation at the 8th IFAC/IFORS Symposium on Identification and System Parameter Estimation, August 27–31, 1988 at Beijing, China

[42] Mehrotra K, Mahapatra PR. A jerk model for tracking highly maneuvering targets. IEEE Transactions on Aerospace and Electronic Systems. 1997;**33**(4):1094-1105

[43] Maybeck PS. Stochastic Models, Estimation, and Control. Vol. 1. New York: Academic Press; 1979

[44] Candy JV. Signal Processing the Model Based Approach. McGraw Hill; 1986

[45] Gemson RMO. Estimation of aircraft aerodynamic derivatives accounting for measurement and process noise By EKF through adaptive filter tuning. [PhD thesis]. Department of Aerospace Engineering, Indian Institute of Science, Bangalore; 1991

[46] Gemson RMO, Ananthasayanam MR. Importance of initial state covariance matrix for the parameter estimation using adaptive extended Kalman filter. In: Proceedings of AIAA Atmospheric Flight Mechanics; 1998. AIAA-98-4153. pp. 94-104

[47] Sarkar AK, Ananthasayanam MR, Vathsal S. Sensitivity of initial state error covariance matrix in a practical adaptive EKF. In: Proceedings of AIAA Atmospheric Flight Mechanics Conference and Exhibit; 6–9 August 2001; Montreal, Canada. AIAA-2001-4202

[48] Shumway RH, Stoffer DS. Time Series Analysis and its Applications. New York: Springer Verlag; 2000

[49] McLachlan GJ, Krishan. The EM Algorithm and Extensions. 2nd ed. John Wiley and Sons; 2008

[50] Bohlin T. Four cases of identification of changing systems. In: System Identification: Advances and Case Studies. 1st ed. Academic Press; 1976

[51] Ljung L. Asymptotic behaviour of the EKF as a parameter estimator for linear systems. IEEE Transactions on Automatic Control. 1979;**24**:36-50

[52] Mehra R. Approaches to adaptive filtering. IEEE Transactions on Automatic Control. 1972;**17**:903-908

[53] Myers KA, Tapley BD. Adaptive sequential estimation with unknown noise statistics. IEEE Transactions on Automatic Control. 1976;**21**:520-525

[54] Ananthasayanam MR, Shyam MM, Naik N, Gemson RMO. A heuristic reference recursive recipe for adaptively tuning the Kalman filter statistics part-1: Formulation and simulation studies. Sadhana. 2016;**41**(12):1473-1490

[55] Shyam MM, Naik N, Gemson RMO, Ananthasayanam MR. A heuristic reference recursive recipe for adaptively tuning the Kalman filter statistics part-2: Real data studies. Sadhana. 2016;**41**(12):1491-1507

[56] Ananthasayanam MR, Suresh HS, Muralidharan MR. GUI Based Software for Teaching Parameter Estimation Technique Using MMLE. Report 2001 FM 1, Dept. of Aerospace Engineering. Bangalore: Indian Institute of Science; 2001

[57] Bavdekar VA, Deshpande AP, Patwardhan SC. Identification of process and measurement noise covariance for state and parameter estimation using extended Kalman filter. Journal of Process Control. 2011;**21**:585-601

[58] Mohamed AH, Schwarz KP. Adaptive Kalman filtering for INS/GPS. Journal of Geodesy. 1999;**73**(4):193-203

[59] Maine RE and Iliff KW. Programmer's Manual for MMLE3, A General Fortran Program for Maximum Likelihood Parameter Estimation. NASA TP-1690; 1981