



# A heuristic reference recursive recipe for adaptively tuning the Kalman filter statistics part-1: formulation and simulation studies

M R ANANTHASAYANAM<sup>1,\*</sup>, M SHYAM MOHAN<sup>2</sup>, NAREN NAIK<sup>3</sup> and R M O GEMSON<sup>4</sup>

<sup>1</sup>Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India

<sup>2</sup>Indian Institute of Technology, Kanpur, India

<sup>3</sup>Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India

<sup>4</sup>Hindustan Aeronautics Limited, Bangalore, India

e-mail: sayanam2005@yahoo.co.in; shyammoh.2014@iitkalumni.org; nnaik@iitk.ac.in; mogratnam@rediffmail.com

MS received 29 April 2015; revised 16 March 2016; accepted 3 June 2016

**Abstract.** Since the innovation of the ubiquitous Kalman filter more than five decades back it is well known that to obtain the best possible estimates the tuning of its statistics  $\mathbf{X}_0$ ,  $\mathbf{P}_0$ ,  $\Theta$ ,  $\mathbf{R}$  and  $\mathbf{Q}$  namely initial state and covariance, unknown parameters, and the measurement and state noise covariances is very crucial. The manual and other approaches have not matured to a routine approach applicable for any general problem. The present reference recursive recipe (RRR) utilizes the prior, posterior, and smoothed state estimates as well as their covariances to balance the state and measurement equations and thus form generalized cost functions. The filter covariance at the end of each pass is heuristically scaled up by the number of data points and further trimmed to provide the  $\mathbf{P}_0$  for subsequent passes. The importance of  $\mathbf{P}_0$  as the probability matching prior between the frequentist approach via optimization and the Bayesian approach of the Kalman filter is stressed. A simultaneous and proper choice for  $\mathbf{Q}$  and  $\mathbf{R}$  based on the filter sample statistics and other covariances leads to a stable filter operation after a few iterations. A typical simulation study of a spring, mass and damper system with a weak nonlinear spring constant by RRR shows it to be better than earlier techniques. Part-2 of the paper further consolidates the present approach based on an analysis of real flight test data.

**Keywords.** Adaptive EKF; expectation maximization; maximum likelihood; covariance matching; recursive parameter estimation; Cramer Rao bound; probability matching prior.

## 1. Introduction

The solution for the linear filtering problem in discrete time was proposed in the famous 1960 paper by Kalman [1]. This was followed for continuous time in 1961 by Kalman and Bucy [2]. Not many know that the enthusiasm that followed soon after Kalman introduced his filter was damped, since the statistics of the process ( $\mathbf{Q}$ ) and measurement noise ( $\mathbf{R}$ ) had to be specified to implement the filter. 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. Kalman when he proposed the filter dealt with only state estimation. 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. 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 add the unknown initial conditions of the state as well. The estimation of the system parameters  $\Theta$ , the initial state  $\mathbf{X}_0$  and its covariance  $\mathbf{P}_0$ , together with  $\mathbf{Q}$  and  $\mathbf{R}$  is called filter tuning. Tuning is mostly done manually even today since the adaptive filtering method has not fully matured in spite of its applications in many fields of science and engineering. Examples are airplane flight test data analysis [3], target tracking [4], evolution of the space debris scenario [5], fusion of GPS and INS data [6], study of the tectonic plate movements [7], high energy physics [8], agriculture, biology and medicine [9], dendroclimatology [10], finance [11], source separation problem in telecommunications, biomedicine, audio, speech and in particular astrophysics [12] and atmospheric data assimilation for weather prediction [13].

In the present work, a reference recursive recipe (RRR) for tuning the Kalman filter is proposed. A simultaneous and proper choice for  $\mathbf{Q}$  and  $\mathbf{R}$  based on the filter sample statistics and certain other covariances leads to a stable filter operation and provides converged results after a few

\*For correspondence

iterations. We have also suggested another statistic for the estimation of  $\mathbf{Q}$  based upon the difference between the stochastic and dynamic trajectories (see section 5.2a). The RRR attains statistical equilibrium after a few filter iterations (without any direct optimization) over the data together with consistency checks. These cost functions (see section 5) help the user to decide on the convergence of RRR and compare its performance with those of other formulations of the Kalman filter. The present RRR is shown to achieve near-optimal Cramer Rao bound (CRB) for the unknown parameters.

In Part-1 of the paper, the problem is set up in an extended Kalman filter (EKF) framework in section 2. The filter tuning and its importance is discussed in section 3. Brief reviews of the earlier adaptive tuning methods and the present approach are discussed in sections 4 and 5, respectively. Section 6 discusses the results of the proposed RRR applied to a simulated spring, mass and damper system with weak nonlinear spring constant. The concluding remarks are made in section 7. Part-2 of the paper shows the effectiveness of the proposed RRR on more involved real airplane flight test data.

## 2. EKF equations

We consider an EKF formulation as it provides the best scenario. Other filter formulations contain the effect of approximations, discretization and other features. Consider the following well-known discrete time nonlinear filtering problem given by

$$x_k = f(x_{k-1}, \Theta, u_{k-1}) + w_k \quad (1)$$

$$Z_k = h(x_k, \Theta) + v_k, \quad k = 1, 2, \dots, N \quad (2)$$

where 'x' is the state vector of size  $n \times 1$ , 'u' is the control input and 'Z' is the measurement vector of size  $m \times 1$ . The 'f' and 'h' are nonlinear functions of state and measurement equations, respectively. The process noise  $w_k \sim \mathcal{N}(0, \mathbf{Q})$  and the measurement noise,  $v_k \sim \mathcal{N}(0, \mathbf{R})$  are assumed to be zero mean additive white Gaussian. The normal or Gaussian distribution is represented by ' $\mathcal{N}$ ' and it is further assumed that

$$E[w_k w_j^T] = \mathbf{Q} \delta(k-j) \text{ with } E[w_k] = 0$$

$$E[v_k v_j^T] = \mathbf{R} \delta(k-j) \text{ with } E[v_k] = 0$$

$$E[w_k v_j^T] = 0 \forall j, \quad k = 1, 2, \dots, N$$

where  $E[\cdot]$  is the expectation operator and  $\delta$  is the Kronecker delta function defined as

$$\delta(k-j) = \begin{cases} 0 & \text{if } k \neq j; \\ 1 & \text{if } k = j. \end{cases}$$

In the EKF formulation the parameter vector ' $\Theta$ ' of size  $p \times 1$  is augmented as additional states

$$\begin{bmatrix} x_k \\ \Theta_k \end{bmatrix} = \begin{bmatrix} f(x_{k-1}, \Theta_{k-1}, u_{k-1}) \\ \Theta_{k-1} \end{bmatrix} + \begin{bmatrix} w_k \\ 0 \end{bmatrix}.$$

The nonlinear filtering problem is now redefined as

$$X_k = f(X_{k-1}) + w_k \quad (3)$$

$$Z_k = h(X_k) + v_k, k = 1, 2, \dots, N \quad (4)$$

where 'X' and 'w' are, respectively, the augmented state and process noise vector of size  $(n+p) \times 1$  and thus  $w_k \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{Q} & 0 \\ 0 & 0 \end{bmatrix}\right)$ . The control input 'u' and 'hat' symbol for estimates are not shown for brevity. A solution to the above problem is summarized in Brown and Hwang [14],

**Initialization:**  $\mathbf{X}_0 = E[X_{t0}],$   
 $\mathbf{P}_0 = E[(\mathbf{X}_0 - X_{t0})(\mathbf{X}_0 - X_{t0})^T]$

**Prediction step:**  $X_{k|k-1} = f(X_{k-1|k-1}),$   
 $\mathbf{P}_{k|k-1} = F_{k-1} \mathbf{P}_{k-1|k-1} F_{k-1}^T + \mathbf{Q}$

**Update step:**  $K_k = \mathbf{P}_{k|k-1} H_k^T (H_k \mathbf{P}_{k|k-1} H_k^T + \mathbf{R})^{-1}$   
 $X_{k|k} = X_{k|k-1} + K_k (Z_k - h(X_{k|k-1})),$   
 $\mathbf{P}_{k|k} = (\mathbf{I} - K_k H_k) \mathbf{P}_{k|k-1}$

where all the symbols have their usual meaning and

true initial state:  $X_{t0}$

initial state estimate :  $X_{0|0} = \mathbf{X}_0$

initial state covariance matrix:  $\mathbf{P}_{0|0} = \mathbf{P}_0$

state Jacobian matrix:  $F_{k-1} = \left[ \frac{\partial f}{\partial X} \right]_{X=X_{k-1|k-1}}$

measurement Jacobian matrix:  $H_k = \left[ \frac{\partial h}{\partial X} \right]_{X=X_{k|k-1}}.$

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. These steps that statistically combine two estimates at any given time point, one from state  $X_{k|k-1}$  and the other from measurement  $Z_k$  equation, are formal only if their uncertainties denoted by their covariances are available. The states can be estimated given the initial  $\mathbf{X}_0$  and  $\mathbf{P}_0$  along with the process noise input with covariance  $\mathbf{Q}$  and updated with measurements with noise covariance  $\mathbf{R}$ . Over a time span in order to match or minimize the difference  $v_k = Z_k - h(X_{k|k-1})$  of the estimates from the states and measurements 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 [15] which is operationally equivalent to minimizing the cost function

$$\begin{aligned} \mathbf{J} &= \frac{1}{N} \sum \mathbf{v}_k (H_k \mathbf{P}_{k|k-1} H_k^T + \mathbf{R})^{-1} \mathbf{v}_k^T \\ &= \mathbf{J}(\mathbf{X}_0, \mathbf{P}_0, \mathbf{Q}, \mathbf{R}, \Theta) \text{ or} \\ &= \mathbf{J}(\mathbf{X}_0, \Theta, \mathbf{K}(\text{traded for } \mathbf{P}_0, \mathbf{Q}, \mathbf{R})) \end{aligned}$$

based on the summation over all the  $N$  measurements and thus solving for either  $\mathbf{X}_0, \mathbf{P}_0, \mathbf{Q}, \mathbf{R}, \Theta$  or solving for  $\mathbf{X}_0, \Theta, \mathbf{K}$  as the case may be. When  $\mathbf{Q} = 0$ , the MMLE is called as the output error method with the Kalman gain matrix being zero. In the usual Kalman filter implementation generally one does not solve for the statistics  $\mathbf{P}_0, \mathbf{Q}$  and  $\mathbf{R}$  but they are adjusted manually to obtain acceptable results. The numerical effort of minimizing  $\mathbf{J}$  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  $\mathbf{X}_0, \mathbf{P}_0, \Theta, \mathbf{R}$  and  $\mathbf{Q}$ . If the above values are not chosen properly then the filter results can be inferior to those from simpler techniques.

### 3. Tuning of the Kalman filter statistics

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! The ghost of filter tuning chases every variant and formulations of the filter. If not tuned properly, it is difficult to infer if the performance of the filter is due to its formulation or filter tuning! Generally the tuning is manual or with ad hoc quick fix solutions such as limiting  $\mathbf{P}$  from going to zero, or adding  $\mathbf{Q}$  to increase  $\mathbf{P}$  before calculating the gain and multiplying  $\mathbf{P}$  by a factor to limit  $\mathbf{K}$ , 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 follows a heuristic approach together with consistency check.

#### 3.1 Qualitative features of the filter statistics

The  $\mathbf{P}_0$  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  $\mathbf{R}$  can be determined from the measured data. The  $\mathbf{Q}$  can account for inaccuracy in the initial conditions and system model, control input and even account for computational errors in the filter. Though  $\mathbf{Q}$  is considered notorious it is the life line of the Kalman filter to do good work. 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 reentry objects. These are handled by using the kinematic relations among the position, velocity, acceleration and even jerk [16] driven by white Gaussian noise  $\mathbf{Q}$  to enable the filter to track these systems. The process noise inhibits the onset of instability of the filter operation.

#### 3.2 Choice of $\mathbf{X}_0$ and $\mathbf{P}_0$ 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  $\mathbf{X}_0$  for the state. The initial parameters values can be guessed if experimental or computational results are available.

The  $\mathbf{P}_0$  is one of the important tuning parameters as stressed by very few like Maybeck [17], Candy [18], and Gemson [19] but most people treat it casually. Generally a guess  $\mathbf{P}_0$  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 work, a proper  $\mathbf{P}_0$  without any  $\mathbf{Q}$  is shown to be possible for the above. The choice of  $\mathbf{P}_0$  can affect the final covariance ( $\mathbf{P}_{N|N}$ ) 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.

#### 3.3 Tuning filter statistics with both $\mathbf{R}$ and $\mathbf{Q}$

When the data contain the effect of both  $\mathbf{R}$  and  $\mathbf{Q}$  it becomes notorious for analysis. The  $\mathbf{R}$  makes the dynamics blurred but  $\mathbf{Q}$  makes the dynamics wander randomly. The interesting point is the filter by tracking the drifted dynamical behaviour even with large  $\mathbf{Q}$ , 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 [20]. 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 why Gemson [19] and Gemson and Ananthasayanam [21] in their approach had to update  $\mathbf{R}$  and  $\mathbf{Q}$  alternately.

When process noise is also present in the simulated data, without solving the involved optimization problem, we look for consistency based on a comparison between the injected and estimated sequences of  $\mathbf{R}$  and  $\mathbf{Q}$ . Further the various cost functions proposed in our work based on balancing the state and measurement equations helped obtain

confidence in the results. When the filter operates through the data it generates prior, posterior, and smoothed state estimates and their covariances, which help generate candidate 'statistic' to estimate both  $\mathbf{R}$  and  $\mathbf{Q}$ .

#### 4. Review of earlier adaptive Kalman filtering approaches

In EKF if the unknown noise covariances are incorrectly specified biased estimates can arise [22, 23]. Even when the system parameters are known, if an inaccurate description of the noise statistics are used the filter may give poor estimates, or even diverge.

We now briefly discuss the four broad adaptive filtering techniques namely the Bayesian, maximum likelihood (ML), covariance matching and correlation as well as a few other approaches.

Every update in a Kalman filter is obviously a Bayesian update. A typical approach of Alspach [24] deals with a bank of autonomous Kalman filter run with a range of Kalman gains. Each one stores a running sum of the square of the residuals. Subsequently it is possible to obtain the estimates of the unknowns based on a weighted sum over the grid points of the gain.

The ML method [20, 25] maximizes the likelihood function containing the unknown covariances  $\mathbf{Q}$  and  $\mathbf{R}$ . Usually time consuming gradient based numerical optimization procedures are required to estimate the unknowns. Shumway and Stoffer [26] proposed an iterative expectation maximization (EM) technique. Here firstly the states are estimated using an initial guess of the unknowns based on a Kalman smoother. Secondly, the unknowns are next updated in the maximization step. This process is repeated till convergence. Bavdekar *et al* [27] used both a direct optimization method and an extended EM method for EKF.

In the covariance matching, the sample filter statistics should be internally consistent with their ensemble properties also provided by the filter. The approaches of Myers and Tapley [28], Gemson [19], Mohamed and Schwarz [29] and Bavdekar *et al* [27] and the present approach are examples for the covariance matching. Myers and Tapley [28] approach using the innovation can at times make the  $\mathbf{R}$  lose its positive definiteness. To overcome this Mohamed and Schwarz [29] have suggested a more stable statistic based on the filtered residue for estimating  $\mathbf{R}$ . Similarly other available statistics are used to estimate  $\mathbf{Q}$ .

The correlation technique depends on the innovation theorem of Kailath [15], which states that the innovation sequence is zero-mean white Gaussian. If the filter is sub-optimal then the innovation mean will be nonzero and its autocorrelation will not follow the Kronecker delta function, indicating that the choices of any or all of the system matrices as well as the covariances are incorrect. This

technique was pioneered by Mehra [31, 32], Carew and Belanger [30] and Belanger [33]. Starting from an assumed value for the unknown  $\mathbf{R}$  and  $\mathbf{Q}$  [31] an initial estimate for the steady state Kalman gain is obtained. The innovations are checked to see if the particular Kalman gain generates a statistically acceptable white noise sequence. But it should be noted that the Kalman gain can take correct value even when  $\mathbf{R}$  and  $\mathbf{Q}$  are incorrect since different combinations of  $\mathbf{R}$  and  $\mathbf{Q}$  can lead to the same gain.

Neethling and Young [34] noted large  $\mathbf{R}$  and  $\mathbf{Q}$  from Mehra's approach. Odelson *et al* [35] showed, based on counter examples, that the mathematical conditions regarding the system and measurement matrices are not sufficient and not necessary in Mehra's [31] work. They also showed that the variance estimates of Mehra are larger and at times negative. Their constrained autocovariance least squares method provided better results and none negative definite.

Valappil and Georgakis [36] assumed that  $\mathbf{R}$  is available and Monte Carlo simulations were run with uncertain parameter values sampled from the assumed normal distribution. Then the difference between the nominal and the random state trajectory over many simulations provide  $\mathbf{Q}$  at any time instant. Manika *et al* [37] identified the critical  $\mathbf{Q}$  by forming two metrics based on the innovation covariance for chosen suitable values of  $\mathbf{X}_0$ ,  $\mathbf{P}_0$  and  $\mathbf{R}$ . These vary from zero to the number of measurements and *vice versa* as  $\mathbf{Q}$  changes from zero to infinity and they proposed that  $\mathbf{Q}$  can be chosen around the intersection point of these two metrics.

Some attempts have been made like Powell [38] using the simplex method, Oshman and Shaviv [39] using the genetic algorithm and controlled random search by Anilkumar [40]. However when the dimension, nonlinearity and the range of search space become large these could become computationally prohibitive and could lead to local minimum. Lau and Lin [41] also discuss the limitations of simulated annealing and particle optimization techniques for filter tuning. One can summarize that deterministic or probabilistic optimization approaches do not appear to be efficient and general for solving the filter tuning problem. We tried if a recursive filtering approach would work and fortuitously it did and will be demonstrated subsequently.

#### 5. RRR for tuning filter statistics

Fundamentally the estimation theory (ET) is an optimization problem. Hence a suitable cost function  $\mathbf{J}$  has to be chosen. Essentially there are two elements in ET: (i) defining a cost function and (ii) adopting a suitable algorithm to minimize the cost function. In general, the log likelihood cost ( $\mathbf{L}$ ) for normally distributed error ( $e$ ) is given by Gemson [19]



$$\mathbf{L}(\Theta) = \frac{1}{N} \sum_{k=1}^N (e_k)^T \mathbf{A}_k^{-1} (e_k) + \log(\det(\mathbf{A}_k)) \quad (5)$$

where  $\mathbf{A}$  is the error covariance matrix and  $\det(\mathbf{A})$  represents determinant of matrix  $\mathbf{A}$ . It may be noted that the parameters ( $\Theta$ ) occur implicitly and not explicitly in the cost function  $\mathbf{L}$ . Since the filter provides many quantities it is possible to form many weighted cost functions based on (i) *a priori* knowledge about the initial estimates, (ii) balancing the measurement equations and (iii) balancing the system equations. One can evolve many cost functions **J0**, **J1**, **J2**, **J3**, **J4**, **J5**, **J6**, **J7** and **J8** as follows:

$$\begin{aligned} \mathbf{J0} &= \frac{1}{2} (\mathbf{X}_0 - \mathbf{X}_{t0})^T \mathbf{P}_0^{-1} (\mathbf{X}_0 - \mathbf{X}_{t0}) \\ \mathbf{J1} &= \frac{1}{N} \sum_{k=1}^N (Z_k - h(X_{k|k-1}))^T S_1^{-1} (Z_k - h(X_{k|k-1})) \\ \mathbf{J2} &= \frac{1}{N} \sum_{k=1}^N (Z_k - h(X_{k|k}))^T S_2^{-1} (Z_k - h(X_{k|k})) \\ \mathbf{J3} &= \frac{1}{N} \sum_{k=1}^N (Z_k - h(X_{k|N}))^T S_3^{-1} (Z_k - h(X_{k|N})) \\ \mathbf{J4} &= \frac{1}{N} \sum_{k=1}^N (Z_k - h(X_{d_{k|N}}))^T (Z_k - h(X_{d_{k|N}})) \\ \mathbf{J5} &= \frac{1}{N} \sum_{k=1}^N (Z_k - h(X_{k|k-1}))^T S_1^{-1} (Z_k - h(X_{k|k-1})) \\ &\quad + \log(\det(S_1)) \\ \mathbf{J6} &= \frac{1}{N} \sum_{k=1}^N w_{1k|N}^T W_{1k}^{-1} w_{1k|N} \\ \mathbf{J7} &= \frac{1}{N} \sum_{k=1}^N w_{2k|N}^T W_{2k}^{-1} w_{2k|N} \\ \mathbf{J8} &= \frac{1}{N} \sum_{k=1}^N w_{3k|k}^T W_{3k}^{-1} w_{3k|k}. \end{aligned}$$

The ‘ $S$ ’ and ‘ $W$ ’ are functions of the second order moments given below [42] and their details are provided in the later sections.

$$\begin{aligned} S_{1k} &= H_k \mathbf{P}_{k|k-1} H_k^T + \mathbf{R} \\ S_{2k} &= -H_{k|k} \mathbf{P}_{k|k} H_{k|k}^T + \mathbf{R} \\ S_{3k} &= -H_{k|N} \mathbf{P}_{k|k-1} H_{k|N}^T + \mathbf{R} \\ W_{1k} &= -\mathbf{P}_{k|N} - F_{k-1|N} \mathbf{P}_{k-1|k-1} F_{k-1|N}^T \\ &\quad + \mathbf{P}_{k,k-1|N} F_{k-1|N}^T + \mathbf{P}_{k,k-1|N}^T F_{k-1|N} + \mathbf{Q} \\ W_{2k} &= -\mathbf{P}_{k|N} - F_{d_{k-1|N}} \mathbf{P}_{k-1|k-1} F_{d_{k-1|N}}^T \\ &\quad + \mathbf{P}_{k,k-1|N} F_{d_{k-1|N}}^T + \mathbf{P}_{k,k-1|N}^T F_{d_{k-1|N}} + \mathbf{Q} \\ W_{3k} &= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k}. \end{aligned}$$

If the initial states are known then **J0** is not necessary but if they are unknown, their estimate and covariance can be obtained, respectively, by the smoothed estimates  $\mathbf{X}_{0|N}$  and  $\mathbf{P}_{0|N}$ . The costs **J1**, **J2** and **J3** for data with measurement noise are expected to tend towards the number of measurements ( $m$ ). The costs **J6**, **J7** and **J8** defined for states with process noise effect are expected to tend towards the number of states ( $n$ ). The cost **J4** is expected to tend towards the trace of  $\mathbf{R}$  for  $\mathbf{Q} = 0$  case and **J5** is the negative log likelihood function. One can formulate any number of cost functions to estimate the parameters and the filter statistics. However it is not possible to estimate the true value of the unknowns but be around them due to statistical fluctuations percolating over all the unknowns.

### 5.1 Choice of $\mathbf{X}_0$ and $\mathbf{P}_0$ in RRR

Commencing from an assumed reasonable initial choice for  $\mathbf{X}_0$ ,  $\mathbf{P}_0$ ,  $\Theta$ ,  $\mathbf{R}$  and  $\mathbf{Q}$  the first filter pass through the data is made. Then a backward smoothing is carried out using the Rauch *et al* [43] smoother. The smoothing leads to the best possible state and parameter estimates and their covariances given all the data. After smoothing the state estimates and their covariances change but not those of the parameters. We next describe how the above are updated for further filter passes through the data to eventually reach statistical equilibrium.

If one uses the smoothed initial state covariance ( $\mathbf{P}_{0|N}$ ) and use it as the  $\mathbf{P}_0$  for the next pass then the final covariance will keep on decreasing with further filter passes and eventually tend towards zero. In order to remedy the above behaviour the final covariance at the end of the pass was scaled up [44] by  $N$  to provide  $\mathbf{P}_0$  at the beginning of the next pass:

$$\mathbf{P}_0 = N \times \mathbf{P}_{N|N}. \quad (6)$$

A heuristic reasoning that can be provided from statistics is that the mean from a sample has an uncertainty  $\mathbf{P}$  that keeps decreasing with sample size as  $\mathbf{P}/N$  where  $\mathbf{P}$  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 method to obtain the  $\mathbf{P}_0$  for the next filter pass. A backward propagation of the final covariance using the estimated parameters did not differ much from the above simple scaled up  $\mathbf{P}_0$ .

The usual recommendation [31] when the states are measured is to set  $\mathbf{P}_0 = \mathbf{R}$ . Even by using the Inverse of Information Matrix (IIM) approach [19] obtained the same estimate for  $\mathbf{P}_0$  as  $\mathbf{R}$ . The IIM is given by

$$\mathbf{P}_0 = \left[ \frac{1}{N} \sum_{k=1}^N F_{k-1}^T H_k^T R^{-1} H_k F_{k-1} \right]^{-1}. \quad (7)$$

After this we have to further trim the above  $\mathbf{P}_0$  to obtain the best possible CRB after some passes. The scaled up  $\mathbf{P}_0$  is a full matrix. Many changes such as using only the diagonal elements and many more variations were tried out. Finally the reference  $\mathbf{P}_0$  to obtain the proper CRB for the parameter estimates turns out to have all the elements are zero (the covariance of all the states and their cross covariance with the parameters as well) except the diagonal elements corresponding to the parameters. If all the elements of the parameter covariances were included and the state and its cross covariances set to zero, it did not make much of a difference in the final results.

**5.1a Probability matching prior interpretation for  $\mathbf{P}_0$ :** If the deterministic Newton–Raphson (NR) optimization approach [45] is considered to provide frequentist results then the Kalman filtering approach corresponds to the Bayesian route. The choice of appropriate probability distribution for  $\mathbf{P}_0$  is the probability matching prior (PMP) that provides a bridge between the above approaches [46]. Consider the simple case of a constant signal with noise. In the frequentist approach the calculation of the mean and standard deviation is direct. However in the Bayesian approach the above result is not reachable unless a proper  $\mathbf{P}_0$  is chosen, which is the PMP.

## 5.2 Estimation of $\mathbf{R}$ and $\mathbf{Q}$ using the EM method

The measurement noise covariance  $\mathbf{R}$  can be estimated using the expectation maximization (EM) algorithm extended to a nonlinear system given by [27]

$$\mathbf{R} = \frac{1}{N} \sum_{k=1}^N E[v_k v_k^T | Z_N]. \quad (8)$$

Consider the measurement noise  $v_k = Z_k - h(X_k)$ , which can be approximated using first order Taylor series expansion around the smoothed estimate  $X_{k|N}$ , given by

$$v_k \approx Z_k - h(X_{k|N}) - H_{k|N} \tilde{X}_{k|N}$$

where  $H_{k|N} = \frac{\partial h}{\partial X}|_{X=X_{k|N}}$  and  $\tilde{X}_{k|N} = X_k - X_{k|N}$

$$\begin{aligned} v_k v_k^T &= Z_k Z_k^T - Z_k h^T(X_{k|N}) - Z_k \tilde{X}_{k|N}^T H_{k|N}^T - h(X_{k|N}) Z_k^T \\ &\quad + h(X_{k|N}) h^T(X_{k|N}) + h(X_{k|N}) \tilde{X}_{k|N}^T H_{k|N}^T - H_{k|N} \tilde{X}_{k|N} Z_k^T \\ &\quad + H_{k|N} \tilde{X}_{k|N} h^T(X_{k|N}) + H_{k|N} \tilde{X}_{k|N} \tilde{X}_{k|N}^T H_{k|N}^T. \end{aligned}$$

We know that

$$\begin{aligned} E[\tilde{X}_{k|N}] &= E[X_k - X_{k|N}] = 0 \\ E[\tilde{X}_{k|N} \tilde{X}_{k|N}^T] &= E[(X_k - X_{k|N})(X_k - X_{k|N})^T] = \mathbf{P}_{k|N}. \end{aligned}$$

Thus the conditional expectation for  $\mathbf{R}$  is given by

$$\begin{aligned} E[v_k v_k^T | Z_N] &= Z_k Z_k^T - Z_k h^T(X_{k|N}) - h(X_{k|N}) Z_k^T \\ &\quad + h(X_{k|N}) h^T(X_{k|N}) + H_{k|N} \mathbf{P}_{k|N} H_{k|N}^T. \end{aligned}$$

Rearranging the above terms and using Eq. (8), we get

$$\mathbf{R} = \frac{1}{N} \sum_{k=1}^N \left\{ (Z_k - h(X_{k|N}))(Z_k - h(X_{k|N}))^T + H_{k|N} \mathbf{P}_{k|N} H_{k|N}^T \right\}. \quad (9)$$

Similarly the process noise covariance  $\mathbf{Q}$  can be estimated as [27]

$$\mathbf{Q} = \frac{1}{N} \sum_{k=1}^N E[w_k w_k^T | Z_N], \quad (10)$$

It results in the following expression for  $\mathbf{Q}$ ;

$$\begin{aligned} \mathbf{Q} &= \frac{1}{N} \sum_{k=1}^N \{ w_{k|N} w_{k|N}^T + \mathbf{P}_{k|N} + F_{k-1|N} \mathbf{P}_{k-1|N} F_{k-1|N}^T \\ &\quad - \mathbf{P}_{k,k-1|N} F_{k-1|N}^T - \mathbf{P}_{k,k-1|N}^T F_{k-1|N} \} \end{aligned} \quad (11)$$

where  $w_{k|N} = X_{k|N} - f(X_{k-1|N})$ ,  $F_{k-1|N} = \left[ \frac{\partial f}{\partial X} \right]_{X=X_{k-1|N}}$ ,  $\mathbf{P}_{k|N}$  is the smoothed covariance,  $\mathbf{P}_{k,k-1|N}$  is the lag-one covariance for  $k = N-1, N-2, \dots, 1$  given by

$$\begin{aligned} \mathbf{P}_{k,k-1|N} &= E[(X_k - X_{k|N})(X_{k-1} - X_{k-1|N})^T] \\ \mathbf{P}_{k,k-1|N} &= \mathbf{P}_{k|k} K_{k-1|N}^T + K_{k|N} (\mathbf{P}_{k+1,k|N} - F_k \mathbf{P}_{k|k}) K_{k-1|N}^T \end{aligned}$$

where  $\mathbf{P}_{N,N-1|N} = (I - K_N H_N) F_{N-1} \mathbf{P}_{N-1|N-1}$ ,  $K_{k|N}$  is the smoothed gain (RTS 1965).

**5.2a The proposed DSDT method for estimating  $\mathbf{Q}$ :** We now estimate  $\mathbf{Q}$  using the difference between the stochastic and dynamical trajectory (DSDT). The stochastic trajectory with the process noise can be approximated using the first order Taylor series expansion around a nominal point ( $X_n$ ) as

$$X_k = f(X_{n_{k-1}}) + f'(X_{n_{k-1}})(X_{k-1} - X_{n_{k-1}}) + w_k. \quad (12)$$

The dynamical trajectory ( $X_d$ ) without the process noise is defined as

$$X_d k = f(X_d k-1) = f(X_{n_{k-1}}) + f'(X_{n_{k-1}})(X_d k-1 - X_{n_{k-1}}) \quad (13)$$

where and  $X_{d0} = X_0$ . It is assumed that the nominal point ( $X_n$ ) of both the above trajectories are close to the estimated dynamical trajectory ( $X_{n_k} \approx X_{d_{k|N}}$ ) where  $X_{d_{k|N}} = f(X_{d_{k-1|N}})$  and  $X_{d0|N} = X_{0|N}$ . Subtracting Eq. (13) from Eq. (12) we get

$$\begin{aligned} X_k - X_{d_k} &= f'(X_{d_{k-1|N}})(X_{k-1} - X_{d_{k-1|N}} - X_{d_{k-1}} + X_{d_{k-1|N}}) + w_k \\ w_k &= X_k - X_{d_k} - F_{d_{k-1|N}}(X_{k-1} - X_{d_{k-1}}) \end{aligned} \quad (14)$$

where the dynamical state Jacobian,  $F_{d_{k-1|N}} = \frac{\partial f}{\partial X}|_{X=X_{d_{k-1|N}}}$ . Thus

$$\begin{aligned} w_k w_k^T &= X_k X_k^T - X_k X_{d_k}^T - X_k X_{k-1}^T F_{d_{k-1|N}}^T + X_k X_{d_{k-1}}^T F_{d_{k-1|N}}^T \\ &\quad - X_{d_k} X_k^T + X_{d_k} X_{d_k}^T + X_{d_k} X_{k-1}^T F_{d_{k-1|N}}^T - X_{d_k} X_{d_{k-1}}^T F_{d_{k-1|N}}^T \\ &\quad - F_{d_{k-1|N}} X_{k-1} X_k^T + F_{d_{k-1|N}} X_{k-1} X_{d_k}^T \\ &\quad + F_{d_{k-1|N}} X_{k-1} X_{k-1}^T F_{d_{k-1|N}}^T - F_{d_{k-1|N}} X_{k-1} X_{d_k}^T F_{d_{k-1|N}}^T \\ &\quad + F_{d_{k-1|N}} X_{d_k} X_{k-1}^T - F_{d_{k-1|N}} X_{d_k} X_{d_{k-1}}^T \\ &\quad - F_{d_{k-1|N}} X_{d_k} X_{k-1}^T F_{d_{k-1|N}}^T + F_{d_{k-1|N}} X_{d_k} X_{d_{k-1}}^T F_{d_{k-1|N}}^T. \end{aligned}$$

We have the following results:

$$\left. \begin{aligned} E[X_k X_k^T | Z_N] &= X_{k|N} X_{k|N}^T + P_{k|N} \\ E[X_k X_{k-1}^T | Z_N] &= X_{k|N} X_{k-1|N}^T + P_{k,k-1|N} \\ E[X_k X_{d_k}^T | Z_N] &= E[X_k | Z_N] E[X_{d_k}^T | Z_N] = X_{k|N} X_{d_k|N}^T \\ E[X_{d_k} X_{d_k}^T | Z_N] &= X_{d_k|N} X_{d_k|N}^T + P_{d_{k|N}} \\ E[X_{d_k} X_{d_{k-1}}^T | Z_N] &= X_{d_k|N} X_{d_{k-1|N}}^T + P_{d_{k,k-1|N}} \end{aligned} \right\} \quad (15)$$

where  $X_{d_{k|N}} = f(X_{d_{k-1|N}})$  is the predicted dynamical state trajectory without the measurement and process noise using the estimated parameter  $\Theta_{N|N}$ . Using Eq. (15) we get

$$\begin{aligned} E[w_k w_k^T | Z_N] &= w_{2k|N} w_{2k|N}^T + P_{k|N} + F_{d_{k-1|N}} P_{k-1|N} F_{d_{k-1|N}}^T \\ &\quad - P_{k,k-1|N} F_{d_{k-1|N}}^T - F_{d_{k-1|N}} P_{k,k-1|N}^T \\ &\quad + P_{d_{k|N}} + F_{d_{k-1|N}} P_{d_{k-1|N}} F_{d_{k-1|N}}^T \\ &\quad - P_{d_{k,k-1|N}} F_{d_{k-1|N}}^T - F_{d_{k-1|N}} P_{d_{k,k-1|N}}^T \end{aligned} \quad (16)$$

where  $w_{2k|N} = X_{k|N} - X_{d_{k|N}} - F_{d_{k-1|N}}(X_{k-1|N} - X_{d_{k-1|N}})$ . Consider the following term;

$$\begin{aligned} X_{d_k} - X_{d_{k|N}} &= f(X_{d_{k-1}}) - f(X_{d_{k-1|N}}) \\ &\approx f(X_{d_{k-1|N}}) + F_{d_{k-1|N}}(X_{d_{k-1}} - X_{d_{k-1|N}}) - f(X_{d_{k-1|N}}) \\ &\approx F_{d_{k-1|N}}(X_{d_{k-1}} - X_{d_{k-1|N}}), \end{aligned} \quad (17)$$

Using Eq. (17), we get the covariance of the dynamical trajectory as

$$\begin{aligned} P_{d_{k|N}} &= E[(X_{d_k} - X_{d_{k|N}})(X_{d_k} - X_{d_{k|N}})^T] \\ &= F_{d_{k-1|N}} P_{d_{k-1|N}} F_{d_{k-1|N}}^T \end{aligned}$$

where  $P_{d_{0|N}} = P_{0|N}$  since  $X_{d_{0|N}} = X_{0|N}$ . The lag-one covariance of dynamical trajectory is

$$\begin{aligned} P_{d_{k,k-1|N}} &= E[(X_{d_k} - X_{d_{k|N}})(X_{d_{k-1}} - X_{d_{k-1|N}})^T] \\ &= F_{d_{k-1|N}} P_{d_{k-1|N}}. \end{aligned}$$

Substituting the value of  $P_{d_{k,k-1|N}}$  and  $P_{d_{k|N}}$  in Eq. (16) and using Eq. (10) we get

$$\begin{aligned} \mathbf{Q} &= \frac{1}{N} \sum_{k=1}^N \{ w_{2k|N} w_{2k|N}^T + P_{k|N} + F_{d_{k-1|N}} P_{k-1|N} F_{d_{k-1|N}}^T \\ &\quad - P_{k,k-1|N} F_{d_{k-1|N}}^T - P_{k,k-1|N}^T F_{d_{k-1|N}} \}. \end{aligned} \quad (18)$$

If  $\mathbf{Q} = 0$  then  $X = X_d$  and assuming that  $P_{0|N} \approx 0$ ,  $\mathbf{R}$  can be estimated as

$$\mathbf{R} \approx \frac{1}{N} \sum_{k=1}^N \{ (Z_k - h(X_{d_{k|N}}))(Z_k - h(X_{d_{k|N}}))^T \}. \quad (19)$$

### 5.3 Possible choices of $\mathbf{R}$

The choice of  $\mathbf{R}$  for the next filter pass can utilize one appropriate among the many that are possible. Bavdekar *et al* [27] used the smoothed residue  $Z_k - h(X_{k|N})$  for  $\mathbf{R}$  estimation using the extended EM method given by

$$\mathbf{R} = \frac{1}{N} \sum_{k=1}^N \{ (Z_k - h(X_{k|N}))(Z_k - h(X_{k|N}))^T + H_{k|N} P_{k|N} H_{k|N}^T \}, \quad (20)$$

The choice of Mohamed and Schwarz (MS) for  $\mathbf{R}$  estimation based on the filtered residue is

$$\mathbf{R} = \frac{1}{N} \sum_{k=1}^N \{ (Z_k - h(X_{k|k}))(Z_k - h(X_{k|k}))^T + H_{k|k} P_{k|k} H_{k|k}^T \}. \quad (21)$$

The choice of Myers and Tapley (MT) for  $\mathbf{R}$  estimation based on the innovation is

$$\mathbf{R} = \frac{1}{N} \sum_{k=1}^N \{ (Z_k - h(X_{k|k-1}))(Z_k - h(X_{k|k-1}))^T - H_k P_{k|k-1} H_k^T \}. \quad (22)$$

All the above measurement noise statistics innovations, filtered residue and smoothed residue are assumed to be of zero mean. The smoothed residue is the best statistic for  $\mathbf{R}$  estimation.

### 5.4 Possible choices of $\mathbf{Q}$

The choice of  $\mathbf{Q}$  for the next filter pass can utilize one appropriate among the many that are possible. Bavdekar *et al* [27] used the smoothed statistic  $X_{k|N} - f(X_{k-1|N})$  for the  $\mathbf{Q}$  estimation using the extended EM method given by

$$\mathbf{Q} = \frac{1}{N} \sum_{k=1}^N \left\{ w1_{k|N} w1_{k|N}^T + \mathbf{P}_{k|N} + F_{k-1|N} \mathbf{P}_{k-1|N} F_{k-1|N}^T - \mathbf{P}_{k,k-1|N} F_{k-1|N}^T - \mathbf{P}_{k,k-1|N}^T F_{k-1|N} \right\} \quad (23)$$

where  $w1_{k|N} = X_{k|N} - f(X_{k-1|N})$ .

The new DSDT statistic for  $\mathbf{Q}$  (section 5.2a) is

$$\mathbf{Q} = \frac{1}{N} \sum_{k=1}^N \left\{ w2_{k|N} w2_{k|N}^T + \mathbf{P}_{k|N} + Fd_{k-1|N} \mathbf{P}_{k-1|N} Fd_{k-1|N}^T - \mathbf{P}_{k,k-1|N} Fd_{k-1|N}^T - \mathbf{P}_{k,k-1|N}^T Fd_{k-1|N} \right\} \quad (24)$$

where  $w2_{k|N} = X_{k|N} - Xd_{k|N} - Fd_{k-1|N}(X_{k-1|N} - Xd_{k-1|N})$ . Mohamed and Schwarz (MS) used innovations and gain for estimating  $\mathbf{Q}$  given by

$$\mathbf{Q} = K_N \left\{ \frac{1}{N} \sum_{k=1}^N (Z_k - h(X_{k|k-1}))(Z_k - h(X_{k|k-1}))^T \right\} K_N^T. \quad (25)$$

The choice of Myers and Tapley (MT) for  $\mathbf{Q}$  is  $w3_{k|k} = X_{k|k} - X_{k|k-1}$  and is given by

$$\mathbf{Q} = \frac{1}{N} \sum_{k=1}^N \left\{ w3_{k|k} w3_{k|k}^T - (F_{k-1} \mathbf{P}_{k|k-1} F_{k-1}^T - \mathbf{P}_{k|k}) \right\}. \quad (26)$$

All the process noise samples,  $w1_{k|N}$ ,  $w2_{k|N}$  and  $w3_{k|k}$ , are assumed to be of zero mean. We note that the smoothed statistics  $w1_{k|N}$  and  $w2_{k|N}$  provide very close results and are the best for the  $\mathbf{Q}$  estimation.

### 5.5 Adaptive tuning and the present RRR

The different methods and options for developing RRR, based on sensitivity studies, are

1.  $\mathbf{P}_0$  can be estimated by scale up, IIM or by smoothing ( $\mathbf{P}_{0|N}$ ).
2. Options for  $\mathbf{P}_0$  split as cov([state-S; parameter-P]) are (a) reference Matrix [0,0;0,\*], (b) diagonal matrix [\*;0;0,\*] and (c) full matrix [\*,\*,\*,\*]. The checkmark (\*) represents a nonzero value at the indicated position. The 'cov' (.) represents covariance matrix.
3. The process noise  $\mathbf{Q}$  can be estimated using Eq. (23), (24), (25) or (26).
4. Options for  $\mathbf{Q} = \text{cov}([S;P])$  are (a) reference matrix [\*;0;0,0], (b) diagonal matrix [\*;0;0,\*] and (c) full matrix [\*,\*,\*,\*].
5. The measurement noise  $\mathbf{R}$  can be estimated using Eq. (20), (21) or (22).

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

1. Given the system model and the measurements the first iteration of EKF is carried out with guess values of  $\mathbf{X}_0$ ,  $\mathbf{P}_0$ ,  $\Theta$ ,  $\mathbf{R}$  and  $\mathbf{Q}$ .
2. Run the extended RTS smoother using the filtered data to get the smoothed state estimate  $X_{k|N}$  and the corresponding smoothed covariance  $\mathbf{P}_{k|N}$ .
3. The  $\mathbf{P}_0$  can be estimated by scale up (Eq. (7)), IIM (Eq. (6)) or the smoothed ( $\mathbf{P}_{0|N}$ ), which will have to be scaled up and modified for obtaining proper results.
4. The  $\mathbf{R}$  and  $\mathbf{Q}$  can be estimated by any of the options as discussed in sections 5.3 and 5.4.
5. EKF is run using the updated estimates of  $\mathbf{X}_0$ ,  $\mathbf{P}_0$ ,  $\Theta$ ,  $\mathbf{Q}$  and  $\mathbf{R}$  at the beginning of further iterations until statistical equilibrium is reached.
6. Different cost functions (J1–J8) are checked for convergence.
7. Many simulation runs (say 50) are carried out by varying the injected measurement ( $v$ ) and process noise ( $w$ ) sequences.

For the  $\mathbf{Q} = 0$  case the value of  $\mathbf{Q}$  is set at  $10^{-10}$  or lower for all iterations to help the filter, which would otherwise generate a pseudo- $\mathbf{Q}$ , and then slowly grind it to zero in hundreds of iterations. For the  $\mathbf{Q} > 0$  case if any of the states is known to have  $\mathbf{Q} = 0$  then it can be set at  $10^{-10}$  or lower. For  $\mathbf{Q} = 0$  case one can estimate  $\mathbf{R}$  even by ignoring the second order terms. It is of interest to note that for  $\mathbf{Q} > 0$  case unless the second order terms of the filter output covariance terms are also included in (i) the estimate for  $\mathbf{R}$  and  $\mathbf{Q}$  using the EM option and (ii) the estimate for  $\mathbf{R}$  using the EM together with  $\mathbf{Q}$  using the DSDT option the  $\mathbf{R}$  and  $\mathbf{Q}$  do not converge to the proper value. The different adaptive approaches analysed in the paper are provided in table 1. Based on a comparative study the proposed RRR is as below.

The reference recursive recipe (RRR)

$\mathbf{Q} = 0$	$\mathbf{Q} > 0$
$\mathbf{X}_0$ : Given or $X_{0 N}$	$\mathbf{X}_0$ : Given or $X_{0 N}$
$\Theta$ : $\Theta_{N N}$	$\Theta$ : $\Theta_{N N}$
$\mathbf{P}_0$ : Scaled up and trimmed-[0,0;0,*]	$\mathbf{P}_0$ : Scaled up and trimmed-[0,0;0,*]
$\mathbf{Q}$ : $10^{-10}$ and trimmed-[*,0;0,0]	$\mathbf{Q}$ : EM/DSDT and trimmed-[*,0;0,0]
$\mathbf{R}$ : EM-diag	$\mathbf{R}$ : EM-diag

The  $\mathbf{P}_0$  is a matrix involving both the uncertainty of the initial state and parameter and their cross-covariance. Thus we can split it into four submatrices. The line ' $\mathbf{P}_0$  : scaled up-[0,0;0,\*]' means that  $\mathbf{P}_0$  is estimated using the scaled up method (section 5.1) and the three submatrices are forced to



**Table 1.** Different adaptive methods for comparative analysis.

Method	Options for $\mathbf{P}_0$	Options for $\mathbf{Q}$	Options for $\mathbf{R}$	Remarks
NR [42]	Not applicable	Not applicable	Using $(Z - h(X_\Theta))$	Applicable only for $\mathbf{Q} = 0$ case.
RRR (present)	Scaled up $\mathbf{P}_{N N}$ (Eq. (6)) trimmed to $[0,0;0,*]$	EM algorithm (Eq. (23)) or DSDT (Eq. (24)) trimmed to $[*; 0; 0, 0]$	Smoothed residue $(Z_k - h(X_{k N}))$ Eq. (20)	Statistical equilibrium and CRBs are achieved
IIM (Similar to Gemson's method)	IIM (Eq. (7)) trimmed to $[0,0;0,*]$	MT method (Eq. (26)) trimmed to $[*; 0; 0, 0]$	Innovations $(Z_k - h(X_{k k-1}))$ Eq. (22)	Initial $\mathbf{R}$ should be close to the true value.
Bavdekar <i>et al</i> method	Smoothed $\mathbf{P}_{0 N}$	EM algorithm (Eq. (23)) with full matrix	Smoothed residue $(Z_k - h(X_{k N}))$ Eq. (20)	Cost functions diverge after some iterations. CRBs are not achieved.
MT Method*	Scaled up $\mathbf{P}_{N N}$ (Eq. (6)) trimmed to $[0,0;0,*]$	MT method (Eq. (26)) trimmed to $[*; 0; 0, 0]$	Innovations $(Z_k - h(X_{k k-1}))$ Eq. (22)	Cost function sometimes oscillates
MS Method*	Scaled up $\mathbf{P}_{N N}$ (Eq. (6)) trimmed to $[0,0;0,*]$	MS Method (Eq. (25)) trimmed to $[*; 0; 0, 0]$	Filtered residue $(Z_k - h(X_{k k}))$ Eq. (21)	$\mathbf{R}$ can go to a very high value

\*Since in MS and MT methods the  $\mathbf{P}_0$  is not specified we have suggested the above.

zero except for the last submatrix involving initial parameter uncertainty as indicated by the tick mark '\*'. The scaling refers to the final  $\mathbf{P}_{N|N}$  multiplied by  $N$  (Eq. (6)), the number of time points, and the trimming refers to setting all the elements of the estimated  $\mathbf{P}_0$  matrix to zero except the ones corresponding to the parameters. Similarly trimming is done for  $\mathbf{Q}$  to set the submatrix involving parameter uncertainty to zero.

Over a fairly wide range of initial values the RRR converges to the same final estimates for all the unknowns, thus showing its stability. We call this as a reference and not as a standard since improvements could be made later when such recursive filtering approaches match the solutions provided by optimization techniques like NR or other involved ones including  $\mathbf{Q}$ . The  $\mathbf{X}_0$  in all cases is either given or obtained by the smoother. The smoothed initial estimate can be split as  $X_{0|N} = (x_{0|N}, \Theta_{0|N})^T$  including both state and parameter. The estimated parameter is taken as  $\Theta_{N|N}$  obtained from  $X_{N|N} = [x_{N|N}, \Theta_{N|N}]$  with covariance  $\mathbf{P}_\Theta$  obtained at the end of the final filter pass over the data,  $\mathbf{P}_{N|N} = [\mathbf{P}_{xx}, \mathbf{P}_{x\Theta}; \mathbf{P}_{\Theta x}, \mathbf{P}_\Theta]$ . Should the measurement data change then the converged parameters also change (!), thus showing sensitivity.

The present RRR contains no direct optimization of any cost function. We have purely iterated the filter on the measured data but after every iteration the unknowns  $\mathbf{P}_0$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$  and  $\Theta$  have been updated as mentioned earlier. Such an iterative procedure together with well chosen updates 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 minimizes any suitable cost function. Checking for the values of the various cost functions with their

expected values provides a good indication as to how best the solutions are. In fact it is such a feature that indicated superiority of the present approach in particular in the real flight test data analysis. Such a difference between the optimization and the present iterative approach corresponds to the frequentist and the Bayesian approach. The crucial success of the present approach has been due to the proper choice of  $\mathbf{P}_0$  by scaling and trimming, which turns out to be PMP, and the appropriate choice of the statistics for  $\mathbf{R}$  and  $\mathbf{Q}$  after every iteration. The perfect match of results based on optimization and purely filtering approach (which are procedurally different though both have Kalman filters in them) would form a very interesting topic for further research.

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

The convergence of any technique even in simulation studies is not a guarantee for a proper solution to the problem. Even the simple case of a linear fit to a set of data with many variants tends to different results [42]. Hence we still lean on simulation studies with exact solutions available to the analyst. Hence presently the filter methods have been applied firstly to a very simple spring, mass and damper system with only  $\mathbf{R}$  and  $\mathbf{Q} = 0$ . For such a situation, the NR procedure [45] served as an anchor for tuning the filter statistics to get the closest possible estimates and the CRB. Subsequently when the process noise is included in the system we looked at the consistency between the injected measurement and process noise sequences and

their statistics. Consider the SMD system with weak non-linear spring constant in the continuous time ( $t$ ) state space form given by

$$\begin{aligned}\dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= -\Theta_1 x_1(t) - \Theta_2 x_2(t) - \Theta_3 x_1^3(t)\end{aligned}$$

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  $\Theta = [\Theta_1, \Theta_2, \Theta_3]^T$  with the true values  $\Theta_{true} = (4, 0.4, 0.6)^T$ . The  $\Theta_3$  is a weak parameter since its values do not affect the system dynamics much. The complete state vector  $X = [x_1, x_2, \Theta_1, \Theta_2, \Theta_3]^T$  is of size  $(n + p) \times 1$ . The measurement equation is given by

$$Z_k = HX_k + v_k$$

where  $H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$  is the measurement matrix of size  $m \times (n + p)$  where  $m = n = 2$  and  $p = 3$ . The values of the noise variances are  $\mathbf{R} = \text{diag}(0.001, 0.004)$  and  $\mathbf{Q} = \text{diag}(0.001, 0.002)$  where ‘diag’ is the diagonal operator as used in MATLAB<sup>®</sup>. All the figures are presented for only one simulation run to prevent cluttering. In the SMD system study, the guess value of  $\mathbf{P}_0$  chosen is  $10^{-1}$  for all states, which is assumed to be a diagonal matrix in the first iteration. The guess value of  $\mathbf{Q}$  chosen is  $10^{-1}$  for all states and zero for the augmented parameters. The guess value of  $\mathbf{R}$  chosen is  $2^{-1}$  for all measurement channels. The initial parameters are chosen to be within  $\pm 20\%$  error. A total of  $N = 100$  measurement data are simulated with the time varying from 0 to 10 s in very small steps of  $\delta t = 0.1$  s. For zero process noise case, the maximum number of iterations is set to 20 over  $n_s = 50$  simulations and for nonzero process noise case it is set to 100 over 50 simulations for obtaining generally four digits accuracy (though not necessary) in the results as presented in tables 2 and 3. In the present RRR it was noticed that generally even if the initial state covariance, initial process and measurement noise covariances were varied over a wide range of powers from  $-3$  to  $+3$  together with the initial parameter values being set to zero one can reach the same estimation results for a given data. Such studies show that RRR leads to a nondiverging and consistent filter performance over many simulations and provides better results when compared with earlier approaches. The convergences of the following quantities [42] through the iterations are analysed.

1. The parameter estimates  $\Theta$  and their covariances  $\mathbf{P}_\Theta$ .
2. The noise covariances  $\mathbf{Q}$  and  $\mathbf{R}$ .
3. The state dynamics without  $\mathbf{R}$  and  $\mathbf{Q}$  based on the estimated parameter after the filter pass through the data  $X_d$ , the prior state  $X_-$ , the posterior state  $X_+$ , the smoothed state  $X_s$  and the measurement  $Z$ .

4. The sample innovation, filtered residue and the smoothed residue along with  $\pm\sigma$  bounds.
5. The estimated measurement and process noise samples as well as their autocorrelations.
6. The cost functions (**J1–J8**) after the final convergence.

### 6.1 Discussion of the results

The SMD system was solved using RRR and with other possible choices for the filter statistics. The following results averaged over 50 simulations are tabulated.

- The  $\Theta$  EKF/NR is the ratio of EKF estimated parameter to that by the NR method.
- The  $\Theta$  EKF/true is the ratio of EKF estimated parameter to that of true value.
- The CRB ratio is the ratio of the square root of the parameter covariance ( $\mathbf{P}_\Theta$ ) estimated by EKF to that of the CRB estimated by the NR method.

- Consistency ratio =  $\frac{\left[ \frac{1}{n_s} \sum_{s=1}^{n_s} (\Theta^s - \bar{\Theta})^2 \right]^{\frac{1}{2}}}{\frac{1}{n_s} \sum_{s=1}^{n_s} \sqrt{\mathbf{P}_\Theta^s}}$  where  $n_s$  is the total number of simulations,  $s$  is the simulation number and  $\bar{\Theta}$  is the sample mean of the estimated parameters.
- Spread factor is a measure of percentage spread seen in the parameter estimates given by

$$\text{spread factor} = \left[ \frac{1}{n_s} \sum_{s=1}^{n_s} \sqrt{(\Theta - \Theta^s)^2 + \mathbf{P}_\Theta^s} \right] \times \frac{100}{|\bar{\Theta}|}.$$

- $\mathbf{R}$  ratio EKF/true is the ratio of the EKF estimated  $\mathbf{R}$  to that of the true value of  $\mathbf{R}$ .
- $\mathbf{R}$  ratio EKF/NR is the ratio of the EKF estimated  $\mathbf{R}$  to that estimated by the NR method.
- $\mathbf{Q}$  ratio is the ratio of the EKF estimated  $\mathbf{Q}$  to that of the true  $\mathbf{Q}$ .
- The mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the cost functions (**J1–J8**) are over many simulations ( $n_s = 50$ ).

6.1a *Without process noise ( $\mathbf{Q} = 0$ ):* Table 2 shows the results for the  $\mathbf{Q} = 0$  case. The appropriate reference results are shown in the second set of rows. The other rows present results with different options. One can note that the parameter estimates, the ratio of  $\mathbf{R}$  and the cost functions are fairly comparable for different options. However it is the ratio of the filter estimated CRBs (whose effect is amplified and shown) by the consistency ratio and the spread factor that differ from the corresponding NR estimates. This feature indicates that the reference procedure with its choice of  $\mathbf{P}_0$ ,  $\Theta$ ,  $\mathbf{Q}$  and  $\mathbf{R}$  updates is about the simplest and best among all possible options. Table 2 shows the effect of using simply the smoothed  $\mathbf{P}_0$  without any scaling but using possible options. In this process-

**Table 2.** Sensitivity study : ( $\mathbf{Q} = 0$ ). No. of iterations = 20, no. of simulations = 50.

Study	$\Theta$ ratio EKF/NR	CRB ratio EKF/NR	Consistency ratio-EKF	Consistency ratio-NR	Spread factor EKF	Spread factor NR	$\mathbf{R}$ ratio EKF/NR	$\mu$ of $\mathbf{J1-J5}$	$\sigma$ of $\mathbf{J1-J5}$
$\mathbf{P}_0$ : Scaled up- [0,0;0,*] $\mathbf{Q}$ : EM- [*;0;0,0] $\mathbf{R}$ : EM-diag	Reference adaptive EKF used for $\mathbf{Q} > 0$ case gives extremely slow convergence of $\mathbf{Q}$ taking hundreds of iterations.								
								1.9704	0.0512
$\mathbf{P}_0$ : Scaled up- [0,0;0,*] $\mathbf{Q}$ : $10^{-10}$ - [*;0;0,0] $\mathbf{R}$ : EM-diag	1.0007	1.0048	1.0140	1.0533	0.8128	0.8126		1.9702	0.0512
	1.0000	0.9832	1.3764	1.3138	1.5488	1.5079	1.0139	1.9999	0.0013
	0.9869	0.9830	1.0919	1.0817	14.5838	14.5843	1.0128	0.0048 -10.3911	0.0008 0.2243
								1.9530	0.0570
$\mathbf{P}_0$ : Scaled up- diag $\mathbf{Q}$ : $10^{-10}$ - [*;0;0,0] $\mathbf{R}$ : EM-diag	0.9998	1.2439	1.0518	1.0533	1.0223	0.8126		1.9534	0.0570
	1.0024	1.3877	1.2966	1.3138	2.1729	1.5079	1.0117	1.9998	0.0050
	1.0255	1.7525	1.1761	1.0817	26.6449	14.5843	1.0061	0.0047 -10.3252	0.0008 0.2308
								1.9668	0.0438
$\mathbf{P}_0$ : Scaled up- full $\mathbf{Q}$ : $10^{-10}$ - [*;0;0,0] $\mathbf{R}$ : EM-diag	0.9991	0.9487	1.3638	1.0533	0.8922	0.8126		1.9670	0.0438
	1.0023	1.1580	1.5696	1.3138	2.0313	1.5079	1.0045	1.9999	0.0014
	1.0471	1.0710	1.8936	1.0817	21.6826	14.5843	1.0006	0.0047 -10.4051	0.0008 0.2293
								1.9690	0.0776
$\mathbf{P}_0$ : IIM- [0,0;0,*] $\mathbf{Q}$ : $10^{-10}$ - [*;0;0,0] $\mathbf{R}$ : EM-diag	1.0014	1.0066	1.1296	1.0533	0.8997	0.8126		1.9710	0.0776
	0.9990	0.9942	2.2338	1.3138	2.3360	1.5079	1.0228	1.9998	0.0015
	0.9667	0.9792	1.3654	1.0817	17.3808	14.5843	1.0204	0.0050 -10.2078	0.0009 0.2349
								1.9530	0.0655
$\mathbf{P}_0$ : IIM-diag $\mathbf{Q}$ : $10^{-10}$ - [*;0;0,0] $\mathbf{R}$ : EM-diag	0.9995	1.2489	1.0476	1.0533	1.0205	0.8126		1.9542	0.0655
	1.0023	1.386	1.5883	1.3138	2.4209	1.5079	1.0160	1.9998	0.0049
	1.0375	1.7504	1.1881	1.0817	26.8883	14.5843	1.0102	0.0049 -10.1918	0.0009 0.2403
								1.9540	0.0656
$\mathbf{P}_0$ : IIM-full $\mathbf{Q}$ : $10^{-10}$ - [*;0;0,0] $\mathbf{R}$ : EM-diag	0.9997	1.2493	1.0479	1.0533	1.0185	0.8126		1.9532	0.0656
	1.0025	1.7885	1.6662	1.3138	2.4988	1.5079	1.0178	1.9999	0.0050
	1.0340	1.7501	1.1851	1.0817	26.7966	14.5843	1.0116	0.0050 -10.2001	0.0009 0.2420
								1.9982	0.0009
$\mathbf{P}_0$ : Smoothed- [0,0;0,*] $\mathbf{Q}$ : $10^{-10}$ - [*;0;0,0] $\mathbf{R}$ : EM-diag	1.0006	0.2361	4.2997	1.0533	0.5240	0.8126		1.9982	0.0009
	1.0002	0.2306	5.7348	1.3138	1.0375	1.5079	0.9990	1.9999	0.0001
	0.9911	0.2315	4.5352	1.0817	9.5986	14.5843	0.9997	0.0048 -10.5276	0.0008 0.2240

Table 2 continued

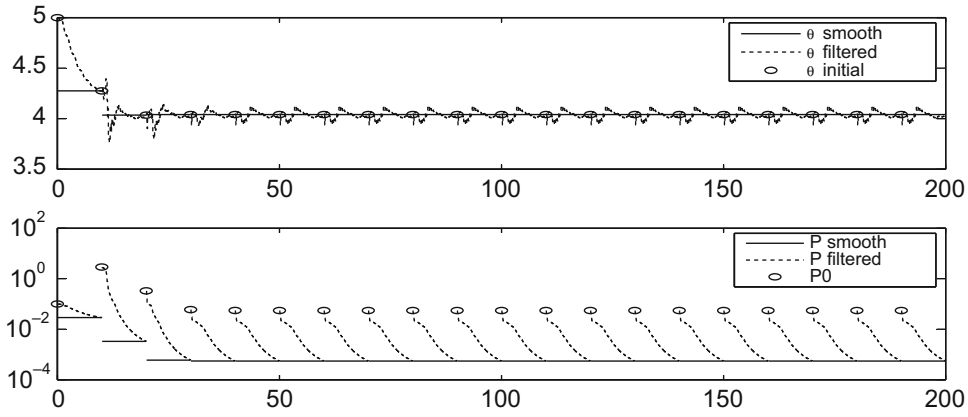
Study	$\Theta$ ratio EKF/NR	CRB ratio EKF/NR	Consistency ratio-EKF	Consistency ratio-NR	Spread factor EKF	Spread factor NR	<b>R</b> ratio EKF/NR	$\mu$ of <b>J1–J5</b>	$\sigma$ of <b>J1–J5</b>
<b>P</b> <sub>0</sub> : Smoothed-diag	1.0021	0.0991	10.8182	1.0533	0.5565	0.8126		1.9960	0.0016
<b>Q</b> : $10^{-10}$ - [*;0;0,0]	0.9999	0.2380	7.4621	1.3138	1.5098	1.5079	0.9904	1.9997	0.0002
<b>R</b> : EM-diag	0.9560	0.0993	16.1755	1.0817	14.2577	14.5843	0.9872	0.0047 –10.5493	0.0008 0.2294
<b>P</b> <sub>0</sub> : Smoothed-full	0.9997	0.3089	4.1863	1.0533	0.6813	0.8126		1.9970 1.9971	0.0011 0.0011
<b>Q</b> : $10^{-10}$ - [*;0;0,0]	1.0011	0.3272	5.4617	1.3138	1.5528	1.5079	0.9891	1.9998	0.0003
<b>R</b> : EM-diag	1.0260	0.4565	4.3984	1.0817	18.0659	14.5843	0.9859	0.0047 –10.5523	0.0008 0.2290

Table 3. Sensitivity study : ( $\mathbf{Q} > 0$ ). No. of iterations = 100, no. of simulations = 50.

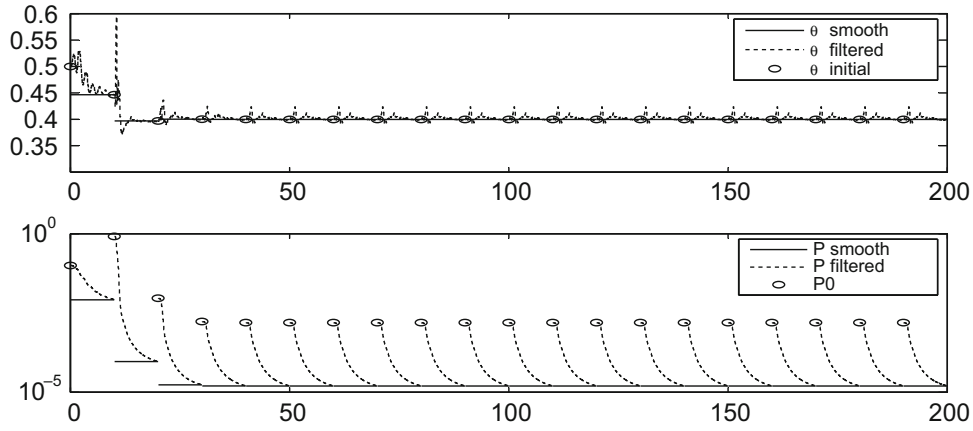
Study	$\Theta$ ratio EKF/ True	Consistency ratio- EKF	Spread factor EKF	<b>R</b> ratio EKF/ True	<b>Q</b> ratio EKF/ True	$\mu$ of <b>J1–J8</b>	$\sigma$ of <b>J1–J8</b>
<b>P</b> <sub>0</sub> : Smoothed-full <b>Q</b> : EM-full <b>R</b> : EM-full	Extended EM algorithm [27] : Cost functions diverges after few iterations. There is a need for a precise stopping condition without which unity ratios cannot be achieved.					1.9650 1.9700 1.9982	0.0224 0.0217 0.0091
<b>P</b> <sub>0</sub> : Scaled up- [0,0;0,*] <b>Q</b> : EM-[*,0;0,0] <b>R</b> : EM-diag	0.9933 1.0201 1.0764	1.0327 1.0772 0.9128	7.9434 23.5655 98.4334	0.9450 0.9135	1.0648 1.0922	0.0709 –8.7886 1.9439 1.9533 1.9585	0.0396 0.2281 0.0422 0.0618 0.0373
<b>P</b> <sub>0</sub> : Smoothed-diag <b>Q</b> : EM-diag <b>R</b> : EM-diag	1.0210 1.0018 0.6806	13.3560 10.0828 13.6464	3.9686 13.6349 51.6473	0.9502 0.9703	1.0286 0.8676	1.9988 1.9996 1.9998 0.0656 –8.9374 2.0031 2.0027 1.9976	0.0241 0.0241 0.0076 0.0325 0.2290 0.0147 0.0239 0.0282

noise-free case, due to the absence of scaling up of the  $\mathbf{P}_0$ , the CRB ratio decreases due to the continuously decreasing smoothed  $\mathbf{P}_{0|N}$  with iterations and leads to incorrect consistency and spread factors in spite of deceptively good **R** ratio and cost function values. The above results show the importance of proper scaling and trimming the  $\mathbf{P}_0$ . When  $\mathbf{Q} = 0$  one can use smoothed residue, filtered residue and innovations with and without second order terms to estimate **R**. In addition another estimate for **R** can be the difference between the measured and predicted dynamics. It turns out that for process noise free case all these options

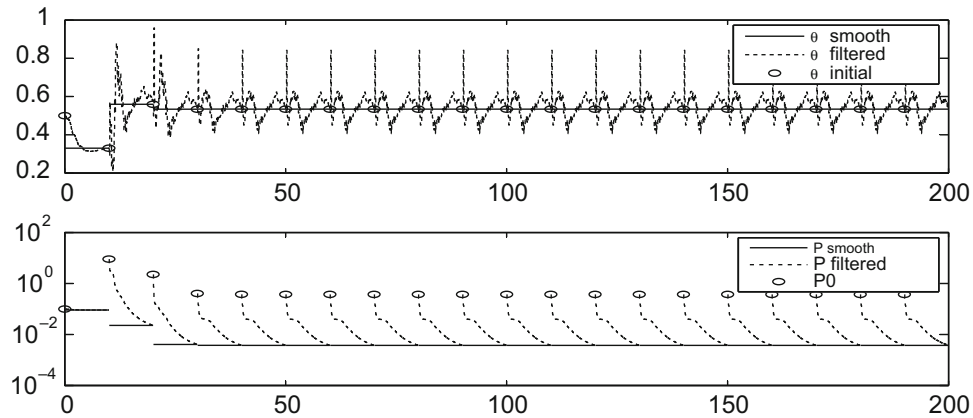
lead statistically to the same results. Figures 1–3 show the variation of the estimated initial parameters and their variances through 20 iterations using the RRR ( $\mathbf{Q} = 0$ ). The  $x$ -axis for the above three plots is the cumulatively increasing time with iterations. For example, in the simulated SMD system, the time instants vary from 0 to 10 s in the first iteration in small time steps of 0.1 seconds. The second iteration has time instants varying from 10 to 20 s, the third iteration with 20–30 s and so on till 200 s for the 20th iteration. The parameter and the uncertainty reach almost their final estimated values with three digits



**Figure 1.** Variation of  $\Theta_1$  and  $P_{\Theta_1}$  with cumulative time and iterations ( $Q = 0$ ).



**Figure 2.** Variation of  $\Theta_2$  and  $P_{\Theta_2}$  with cumulative time and iterations ( $Q = 0$ ).

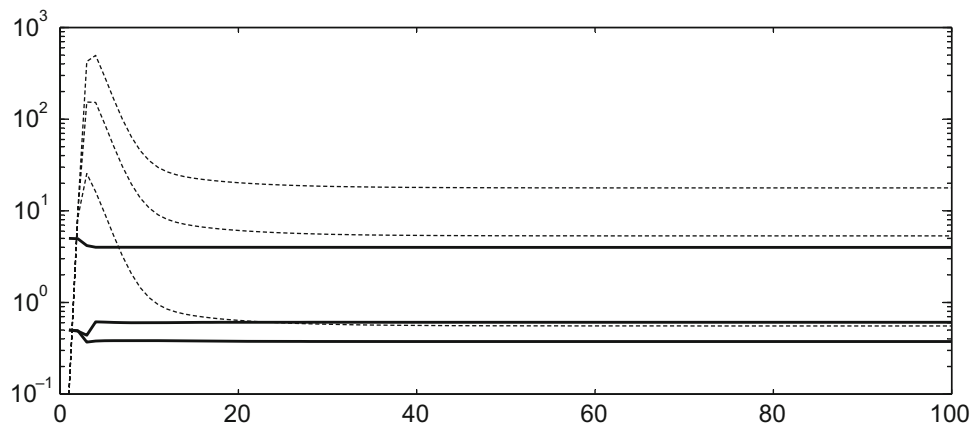


**Figure 3.** Variation of  $\Theta_3$  and  $P_{\Theta_3}$  with cumulative time and iterations ( $Q=0$ ).

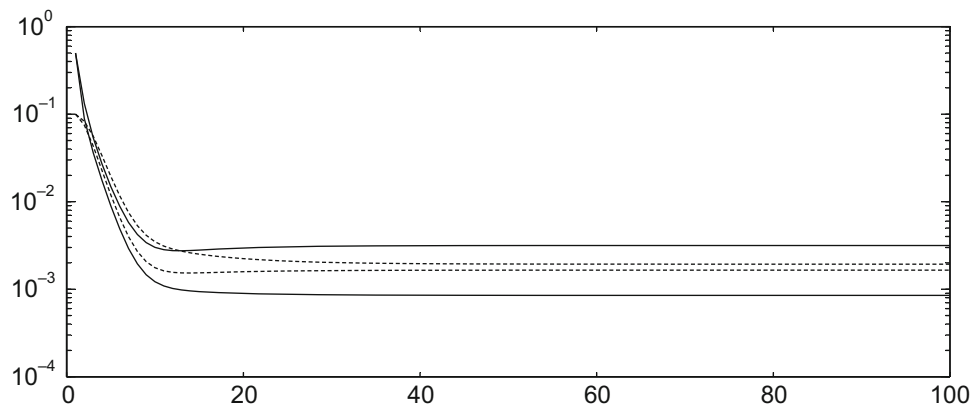
accuracy in about two and five iterations, respectively. The correlation coefficient matrix obtained from the parameter covariance matrix  $P_{\Theta}$  of a typical data set with zero process noise is

$$C = \begin{bmatrix} 1.0000 & -0.0024 & -0.9387 \\ -0.0024 & 1.0000 & 0.1193 \\ -0.9387 & 0.1193 & 1.0000 \end{bmatrix}.$$

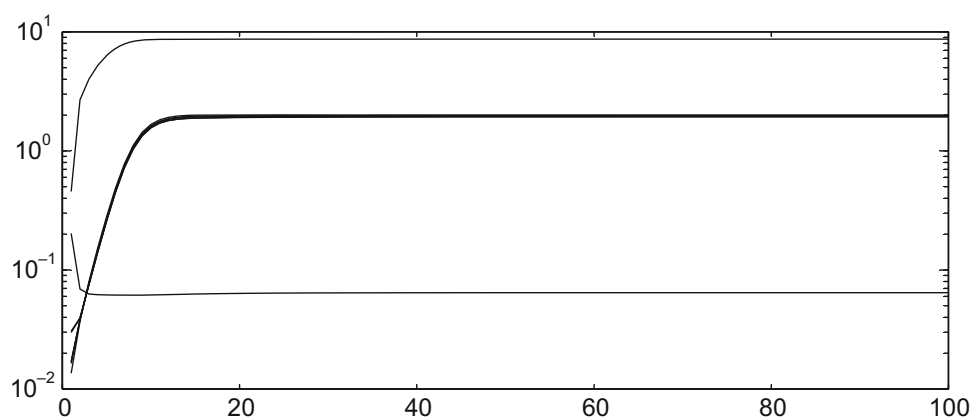




**Figure 4.** Variation of initial parameters  $\Theta_0$ (continuous) and its  $P_0$  (dashed) with iterations.



**Figure 5.** Variation of  $Q$  (dashed) and  $R$  (continuous) with iterations.

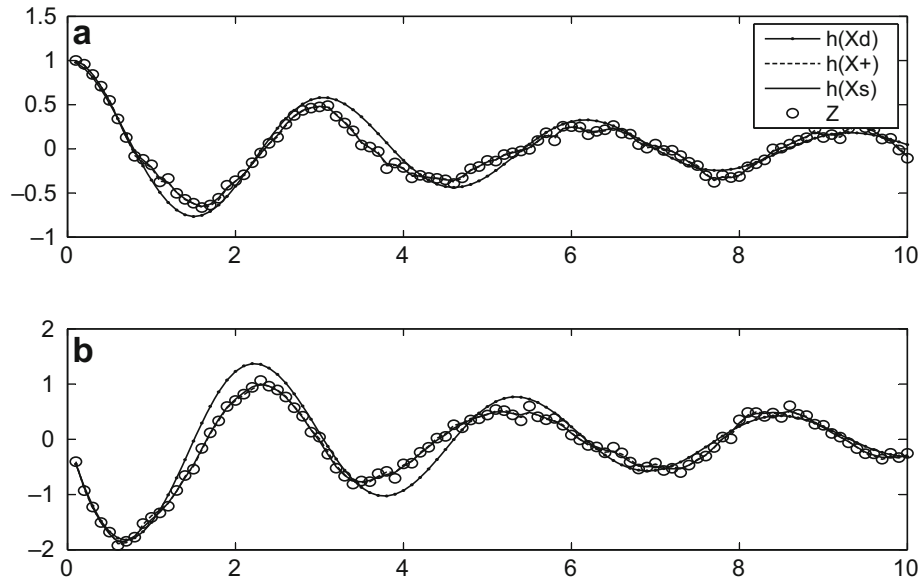


**Figure 6.** Variation of different costs (**J1–J8**) with iterations.

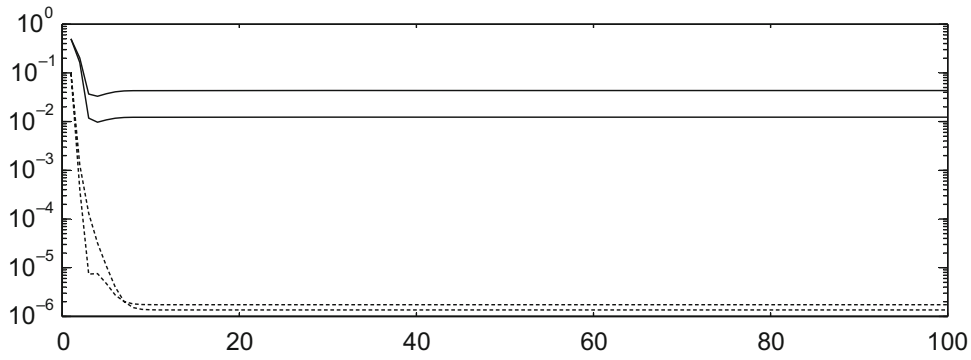
Over a range of displacement, the weak nonlinear spring constant ( $\Theta_3$ ) can be estimated with high correlation only with the linear spring constant ( $\Theta_1$ ). If the range of displacement is increased then the above correlation reduces. However the damping coefficient ( $\Theta_2$ ) is estimated with

modest correlation with other parameters since it is driven by velocity than displacement.

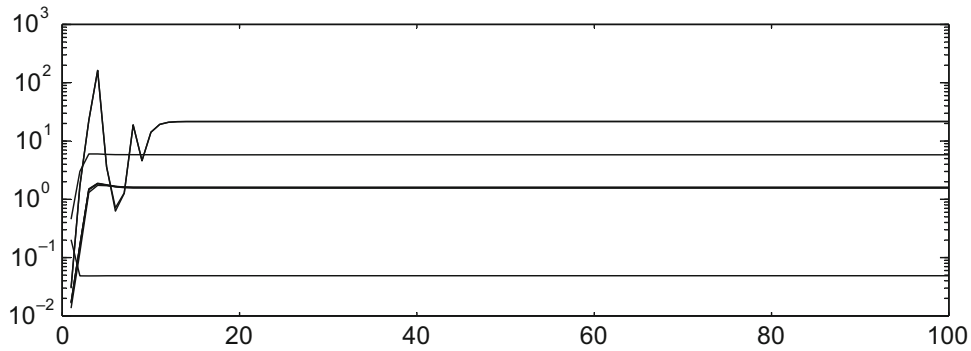
6.1b *With process noise ( $Q > 0$ )*: Figures 4, 5 and 6 show, respectively, the variation of (i) the initially estimated  $\Theta$



**Figure 7.** Comparison of the predicted dynamics  $h(X_d)$ , posterior  $h(X_+)$ , smoothed  $h(X_s)$  and the measurement  $Z$  corresponding to the (a) displacement and (b) velocity from RRR.



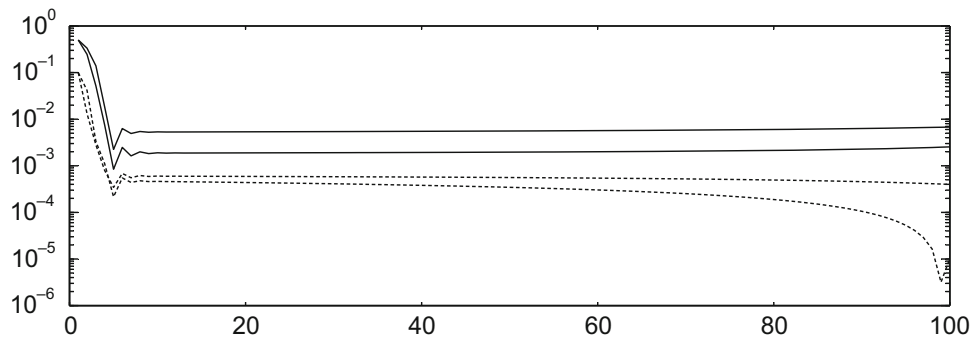
**Figure 8.** Variation of  $Q$  (dashed) and  $R$  (continuous) with iterations using the MS method.



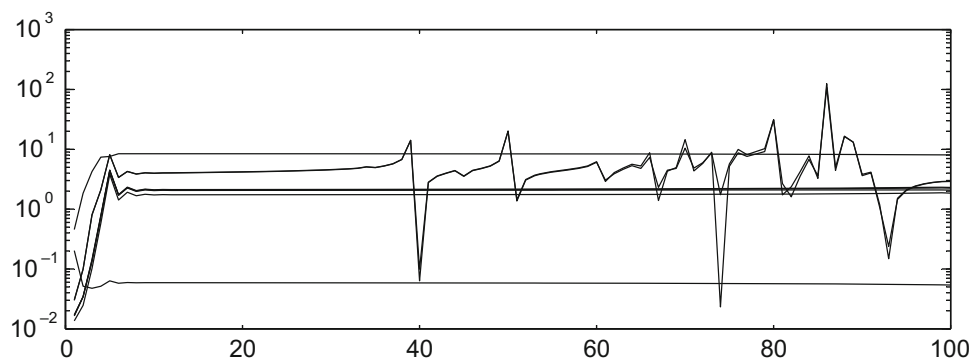
**Figure 9.** Variation of different costs ( $J1$ – $J8$ ) with iterations using the MS method.

and its variance, (ii)  $R$  and  $Q$  and (iii) the different cost functions ( $J1$ – $J8$ ) through 100 iterations for the RRR ( $Q > 0$ ) case. It has been further extensively checked and

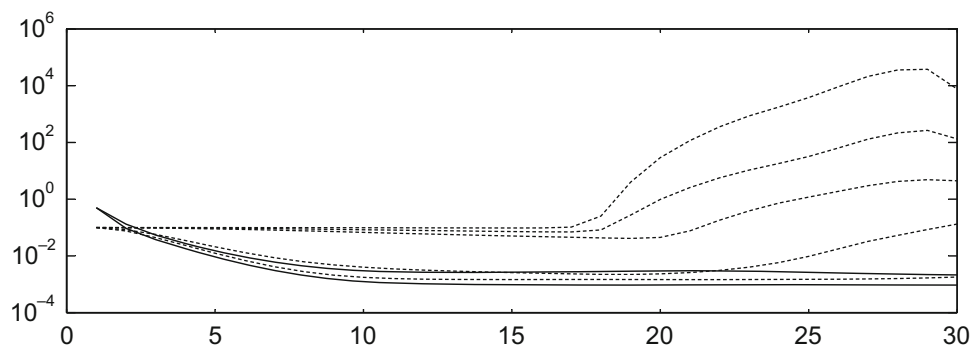
reported by Shyam *et al* [42] that the variations of the sample innovations, filtered residue and smoothed residue are consistent with their  $\pm\sigma$  bound. In the EKF approach



**Figure 10.** Variation of  $\mathbf{Q}$  (dashed) and  $\mathbf{R}$  (continuous) with iterations using the MT method.



**Figure 11.** Variation of different costs ( $\mathbf{J1-J8}$ ) with iterations using the MT method.



**Figure 12.** Variation of  $\mathbf{Q}$  (dashed) and  $\mathbf{R}$  (continuous) with iterations using the [27] method.

since most of the quantities are Gaussian or approximated as quasi-Gaussian and one would expect that all the above quantities are close to being Gaussian and hence around one-third of the total sample points to be outside the  $\pm\sigma$  bound. Similarly the injected and estimated measurement noise distributions during the final iteration were very close to each other, also their autocorrelations are ideally expected to be close to the Kronecker delta function, which provides confidence in the proposed filter algorithm.

Figure 7 shows the predicted dynamics and filtered and smoothed estimate at the last iteration. There is an expected mismatch in the estimated dynamics (without the effect of  $\mathbf{R}$  and  $\mathbf{Q}$ ) and the measurements made on the wandering dynamics, indicating the presence of process noise. Figures 8–13 show the variation of the noise  $\mathbf{R}$  and  $\mathbf{Q}$  estimates and the cost functions  $\mathbf{J1-J8}$  for MS, MT and Bavedkar *et al* approaches. The important feature to be noted in the MS is that though the filter converges it leads

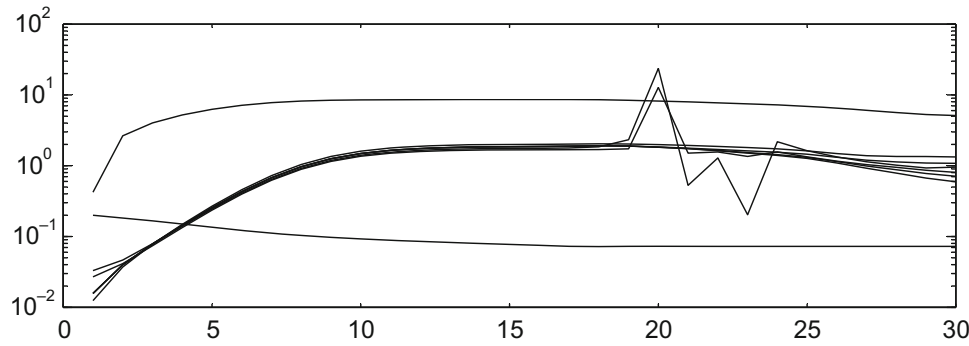


Figure 13. Variation of different costs (J1–J8) with iterations using the [27] method.

to inaccurate estimates for the noise statistics and cost functions. The MT and Bavedkar *et al* approaches in general do not provide systematic variation and convergence of the estimates. Table 3 shows for the reference case in the second row that the cost functions J1–J3 and J6–J8 are close to their expected values. The third row shows, using the smoothed  $\mathbf{P}_0$  without scaling and no stopping condition, that the low spread factor is deceptive since there is no consistency in the parameter estimates with its covariance.

## 7. Conclusions

A comparative study among the existing adaptive techniques suggested by Myers and Tapley, Mohamed and Schwarz and Bavdekar *et al* is carried out and a reference recursive recipe (RRR) for tuning the Kalman filter is proposed. A new statistic for the estimation of  $\mathbf{Q}$  based upon the difference between the stochastic and dynamic trajectories (DSDT) was introduced based on an extended EM method. The different cost functions (J1–J8) help the user to reach deceptive to decisive convergence. The proposed RRR achieves the Cramer Rao bound (CRB) of the unknown parameters and provides statistically equilibrium solution after a few iterations. The importance of  $\mathbf{P}_0$  has been demonstrated and the best possible  $\mathbf{Q}$  and  $\mathbf{R}$  have been obtained.

## Acknowledgements

Our grateful thanks are due to Profs R M Vasu (Department of Instrumentation and Applied Physics), D Roy (Department of Civil Engineering) and M R Muralidharan (Supercomputer Education and Research Centre) for help in a number of ways without which this work would just not have been possible at all and also for providing computational facilities at the IISc, Bangalore.

## References

- [1] Kalman R E 1960 A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* 82 (Series D): 35–45
- [2] Kalman R E, Bucy R S 1961 New results in linear filtering and prediction theory. *J. Basic Eng.* 83(1): 95–108
- [3] Klein V and Morelli E A 2006 Aircraft system identification: theory and practice. In: *AIAA Edu. Series*
- [4] Bar-Shalom Y, Rong Li X and Kirubarajan T 2000 *Estimation with applications to tracking and navigation, theory, algorithm and software*. John Wiley and Sons. Inc
- [5] Ananthasayanam M R, Anilkumar A K and Subba Rao P V 2006 New approach for the evolution and expansion of space debris Scenario. *J. Spacecraft Rockets* 43(6): 1271–1282
- [6] Grewal M S, Lawrence R W and Andrews A P 2007 *Global positioning systems, inertial Navigation, and Integration*, 2nd Edition, John Wiley & Sons, Inc.
- [7] Kleusberg A and Teunissen P J G 1996 *GPS for geodesy*, 1st edition, Springer
- [8] Fruhwirth R, Regier M, Bock R K, Grote H and Notz D 2000 Data analysis techniques for high-energy physics. In: *Cambridge monographs on particle physics, nuclear physics and cosmology*
- [9] Federer W T and Murthy B R 1998 *Kalman filter bibliography : agriculture, biology, and medicine*. Technical Report BU-1436-M, Department of Biometrics, Cornell University
- [10] Visser H and Molenaar J 1988 Kalman filter analysis in dendroclimatology. *Biometrics* 44: 929–940
- [11] Wells C 1996 *The Kalman filter in finance*. Springer-Science+Business Media, BV
- [12] Costagli M and Kuruoglu E E 2007 Image separation using particle filters. *Digital Signal Process.* 17: 935–946
- [13] Evensen G 2009 *Data assimilation: the ensemble Kalman Filter*. Springer Verlag .
- [14] Brown R and Hwang P 2012 *Introduction to random signals and applied Kalman filtering, with MATLAB exercises*. John Wiley and Sons, Inc 4th edition
- [15] Kailath T 1970 An innovation approach to detection and estimation theory. *Proc. IEEE* 58(5): 680–695
- [16] Mehrotra K Mahapatra P R 1997 A jerk model for tracking highly maneuvering targets. *IEEE Trans. Aerospace Electron. Syst.* 33(4)

- [17] Maybeck P S 1979 Stochastic models, estimation, and control. Volume 1, New York: Academic Press
- [18] Candy J V 1986 *Signal processing the Model based approach*. McGraw Hill
- [19] Gemson R M O 1991 *Estimation of aircraft aerodynamic derivatives accounting for measurement and process noise By EKF through adaptive filter tuning*. PhD Thesis, Department of Aerospace Engineering, IISc, Bangalore
- [20] Bohlin T 1976 Four cases of identification of changing systems. In: *System identification: advances and case studies*, 1st edition. Academic Press
- [21] Gemson R M O and Ananthasayanam M R 1998 *Importance of initial state covariance matrix for the parameter estimation using adaptive extended Kalman filter*. AIAA-98-4153, pp. 94–104
- [22] Ljung L 1979 Asymptotic behaviour of the EKF as a parameter estimator for linear systems. *IEEE Trans. Autom. Control* AC 24: 36–50
- [23] Ljungquist D and Balchen J G 1994 Recursive prediction error methods for online estimation in nonlinear state space models. *Model. Identifi. Control* 15(2): 109–121
- [24] Alspach D 1974 A parallel filtering algorithm for linear systems with unknown time varying noise statistics. *IEEE Trans. Autom. Control* 19(5): 552–556
- [25] Kashyap R 1970 Maximum likelihood identification of stochastic linear systems. *IEEE Trans. Auto. Control* 15(1): 25–34
- [26] Shumway R H and Stoffer D S 2000 *Time series analysis and its applications*. Springer, Verlag, NY
- [27] Bavdekar V A, Deshpande A P and Patwardhan S C 2011 Identification of process and measurement noise covariance for state and parameter estimation using extended Kalman filter. *J. Process Control* 21: 585–601
- [28] Myers K A and Tapley B D 1976 Adaptive sequential estimation with unknown noise statistics. *IEEE Trans. Autom. Control* 21: 520–525
- [29] Mohamed A H and Schwarz K P 1999 Adaptive Kalman filtering for INS/GPS. *J. Geodesy* 73(4): 193–203
- [30] Carew B and Belanger P R 1973 Identification of optimum filter steady state gain for systems with unknown noise covariances. *IEEE Trans. Auto. Control* 18(6): 582–587
- [31] Mehra R K 1970 On the identification of variances and adaptive Kalman filtering. *IEEE Trans. Automat. Control* 15(2): 175–184
- [32] Mehra R 1972 Approaches to adaptive filtering. *IEEE Trans. Autom. Control* 17: 903–908
- [33] Belanger P R 1974 Estimation of noise covariances for a linear time-varying stochastic process. *Automatica* 10(3): 267–275
- [34] Neethling C and Young P 1974 Comments on Identification of optimum filter steady state gain for systems with unknown noise covariances. *IEEE Trans. Autom. Control* 19(5): 623–625
- [35] Odelson B J, Lutz A and Rawlings J B 2006 The autocovariance-least squares method for estimating covariances: application to model-based control of chemical reactors. *IEEE Trans. Control Syst. Technol.* 14(3): 532–540
- [36] Valappil J and Georgakis C 2000 Systematic estimation of state noise statistics for extended Kalman filters. *AIChE J.* 46(2): 292–308
- [37] Manika S, Bhaswati G and Ratna G 2014 Robustness and sensitivity metrics for tuning the extended Kalman filter. *IEEE Trans. Instrum. Meas.* 63(4): 964–971
- [38] Powell T D 2002 Automated tuning of an extended Kalman filter using the downhill simplex algorithm. *J. Guidance Control Dyn.* 25(5): 901–908
- [39] Oshman Y and Shaviv I 2000 *Optimal tuning of a Kalman filter using genetic algorithm*. AIAA Paper 2000-4558
- [40] Anilkumar A K 2000 *Application of controlled random search optimisation technique in MMLE with process noise*. MSc Thesis, Department of Aerospace Engineering, IISc, Bangalore
- [41] Lau T and Lin K 2011 Evolutionary tuning of sigma-point Kalman filters. *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 771–776
- [42] Shyam M M, Naren Naik, Gemson R M O and Ananthasayanam M R 2015 *Introduction to the Kalman filter and tuning its statistics for near optimal estimates and Cramer Rao bound*. TR/EE2015/401, IIT Kanpur, <http://arxiv.org/abs/1503.04313>
- [43] Rauch H E, Tung F and Striebel C T 1965 Maximum likelihood estimates of linear dynamic systems. *AIAA J.* 3(8): 1445–1450
- [44] Shyam M M 2014 *An iterative tuning strategy for achieving Cramer Rao bound using extended Kalman filter for a parameter estimation problem*. MTech Thesis, IIT Kanpur
- [45] Ananthasayanam M R, Suresh H S and Muralidharan M R 2001 *GUI based software for teaching parameter estimation technique using MMLE*. Report 2001 FM 1, IISc
- [46] Samaniego F J 2011 *A Comparison of the Bayesian and frequentist approaches to estimation*. Springer Science+Business Media, LLC