Target Tracking Under Communication Constraints

Muhammad Amjad Raza

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Electrical and Electronic Engineering
THE UNIVERSITY OF MELBOURNE

May 2016

Declaration

This is to certify that

1. the thesis comprises only my original work towards the PhD,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Muhammad Amjad Raza, May 2016

Acknowledgements

I would like to thank my supervisor Robin J. Evans for his time, help, support and especially correcting for very naive mistakes . I am truly grateful for my co-supervisor Mark R. Morelande for his contributions towards my thesis. The assistance and mentor-ship of Darko Musicki is unforgettable as my co-supervisor.

A special thanks also goes to Girish Nair for providing me with the opportunity to be involved in demonstrating the undergraduate course (Probability and Random Models). Also, I would like to thank the University of Melbourne and National ICT Australia (NICTA) for providing financial support for my Ph.D. project.

Finally, without the understanding and support from my parents, wife and family, it would have been impossible achieving this milestone.

Dedicated to the memory of Professor Darko Musicki (April 02, 1957 - June 8, 2014)

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Overview

THE subject of this thesis is target tracking and fusion under communication constraints for multi sensor surveillance systems. We estimate the state of the target observed by a sensor by designing a filter best capable of handling uncertainties associated with the dynamics of the target as well as uncertainties associated with sensor measurements. The processing of the collected measurements is carried out at each sensor platform and track state information is transmitted to a fusion center for maintenance of the global track state. The communication channels between sensor platforms and the global processing unit can be wired or wireless. In the literature, this form of tracking is widely referred as distributed tracking. We study distributed tracking where we have restrictions on communication bandwidth availability. In other words, we minimise the communication load while ensuring the tracking efficiency is not degraded below a specified level. The communication between a sensor platform and the fusion center can be two way (in the form of feedback from the fusion center) or just one way (from sensor to the fusion center). Two way communication requires more communication channels leading to increased bandwidth usage. In this work, we consider only one way communication, i.e., sensor transmits only.

The structure of a typical multi sensor network with one central processing unit is as follows: Consider M sensors connected to a fusion center with forward link communications, i.e., sensor platform equipped with transmitter only while receiver resides at the fusion center. In classical centralised tracking, every sensor sends its collected mea-

surements to the fusion center on a dedicated communication channel and all tracking is performed at the fusion center. The benefit of centralised tracking includes optimality, however it consumes huge bandwidth and also demands highly reliable communication channels. On the other hand, local tracking is performed at the sensor platform in distributed tracking. Each sensor is equipped with capabilities to maintain its local track and can send information regarding track state to the fusion center. One of the major benefits of distributed tracking is that local tracking remains intact even if communication links are disrupted. Moreover, distributed tracking also helps to reduce the communication load by providing the choice to transmit less bandwidth demanding data to the fusion center. We adopt a distributed tracking approach to solve the problem of tracking under bandwidth constrained situations.

In this thesis, we study parameter estimation and single target tracking both with and without clutter. The communication requirement increases dramatically in centralised tracking when we need to transmit target as well as clutter measurements to the fusion center. Also, increasing clutter measurement density increases the communication load. By adopting distributed tracking, we have control on what to transmit. Also in distributed tracking, we can choose what track state information to send to the fusion center. The transmission of estimated track state, equivalent innovations and tracklets have been discussed in the literature.

The first contribution to the thesis is distributed tracking when we transmit equivalent measurements (also known as tracklets) to the fusion center. The benefit of transmission of equivalent measurements over equivalent innovations is avoiding random walk phenomenon. We are able to reduce communication load by reducing the dimensionality of equivalent measurements. The compression of equivalent measurements is achieved by a scaling matrix and a compression matrix. We also propose that rather than sending equivalent measurements on every scan, we send them after 2, 3 or N_S scans. This results in a small performance loss when adaptive transmission of equivalent measurements is used. This procedure helps in reducing communication requirements at the cost of a minor loss in tracking performance. This approach is also applicable when clutter is present.

The second part of this thesis discusses distributed tracking in clutter. We assume the tracking environment is heavily cluttered and the sensor collects object detections along with clutter measurements. It is expensive to transmit all received measurements to the fusion center, so we process these measurements at the sensor platform. The standard Probabilistic Data Association Filter (PDAF) is used at the sensor platform. We calculate the scaled sum of all received measurements falling inside the local PDAF gate and transmit the resultant measurement to the fusion center. We propose a Compressed Filter (CF) which operates on the received compressed measurements at the fusion center. Our results show that the tracking performance of the Compressed Filter is promising. By applying the proposed compression at the sensor node and using the proposed CF, we are able to reduce the communication bandwidth while maintaining reasonable tracking performance at the fusion center.

The third part of the thesis is a novel approach to distributed estimation and tracking. We transmit only one (random) sample or particle of the local track posterior distribution at each time stamp to the fusion center. These samples are drawn independently. At the far end, we update the weight of every received particle upto time k and the most recent received particle exhibits highest weight. As we progress in time, the number of received particles grows and our estimate increasingly improves. This idea is similar to importance sampling however the target distribution for each received sample is different at each time. The estimate is a weighted sum of all received samples upto the current time. The bulk of computations at the far end are devoted to updating the weight of each received sample. We study consistency of the above proposed estimator. We show that this estimate is unbiased and the variance of the estimate is bounded.

We cover three distributed estimation/tracking methods and include simulations to support the results.

1.2 Outline of Thesis

In this thesis, we mainly discuss distributed tracking under communication constraints. We cover three distributed estimation/tracking methods and provide analysis and sim-

ulations to support the results. Chapter 2 presents an overview of existing tracking algorithms. We start with linear Bayesian tracking and proceed to non-linear trackers. We discuss non-linear particle filters and certain variants. We also study target tracking algorithms for cluttered environments. Finally, we discuss the use of adaptive importance methods for tracking and parameter estimation.

A detailed discussion on use of equivalent measurements and scaled equivalent measurements is presented in Chapter 3. We overview equivalent innovations and tracklets. We include our proposed new method of dimensionality reduction for equivalent measurement and show optimality via analysis and simulation.

We propose the Compressed Filter for distributed tracking in Chapter 4. We consider tracking in a cluttered environment. The weighted sum of the measurements falling inside a local Probabilistic Data Association (PDA) tracker gate is transmitted to the fusion center and the Compressed Filter is used to extract measurement information from the received information. This work has been published [55].

Chapter 5 proposes a novel distributed tracking algorithm based on transmitting independently drawn samples/particles from the local posterior distribution to the fusion center.

Convergence properties of the proposed estimator and sample weights are discussed in Chapter 6. We also include simulations for the proposed particle based distributed algorithm.

Conclusions and discussions of findings are outlined in Chapter 7. This chapter also includes possible future extensions of the ideas presented in this thesis.

1.3 Major Contributions

We discuss distributed tracking and include three major contributions when low communication bandwidth between a sensor platform and the central processing unit is important.

1. Transmission of compressed/scaled equivalent measurements is studied. Adaptive transmission of this information is explored in order to reduce communication

load. We study several types of compression technique discussed in the literature. We show that when we use the output matrix H as a compression matrix, we achieve optimal tracking results at the fusion center.

2. The second major contribution is to propose and study a Compressed Filter for tracking in clutter. A novel filter is proposed which operates on compressed received measurement information. We compress received measurements at the sensor platform by forming a weighted sum of target and clutter measurements falling inside the tracking gate. We transmit the resultant compressed measurement along with associated weights to the fusion center.
3. The third major contribution is the development of the new distributed estimation algorithm based on transmission of random samples of the local posterior distribution from a sensor platform to the fusion center. We transmit only one sample of the local posterior distribution from the local sensor platform at every scan time and update the weighting of all received samples at the fusion center. We show that the proposed estimator is unbiased and variance of the estimator is bounded. Furthermore, we study the properties of the weights and the estimator. We explore the concept of an adaptive importance density. Our results are developed for a single sensor platform. These results can easily be extended to the multi sensor case where each sensor transmits only one sample from its local posterior probability distribution function to the fusion center. The global tracking is maintained at the fusion center using all samples received at time k and previous times from all sensors in the network.

Chapter 2
Overview of Tracking

This chapter overviews both centralized tracking and distributed tracking. We also discuss tracking with and without clutter. We formulate the problem of tracking and discuss linear and non-linear tracking filters. The traditional Kalman filter and some of its variants for non-linear problems are discussed first. Also, we include recent developments on particle filter type estimation and tracking.

2.1 Introduction

TARGET tracking has been widely studied in the literature including applications in defence, video, vehicle tracking, robot movement tracking, civil aviation and many more. In target tracking, we receive measurements from sensors and estimate parameters of interest based on the received measurements. Received measurements are always noisy and tracking algorithms must mitigate the effect of measurement noise and process uncertainty.

A target tracking system consists of three main subsystems. The first subsystem is the sensor which makes measurements of the surveillance region. There are various kinds of sensors to perform the sensing work. One of the most commonly deployed sensors is Radio Detection and Ranging (RADAR).

A traditional RADAR transmits electromagnetic waves of a certain frequency and beam shape and then records the reflected waves from the surveillance region. Signal processing techniques are then used to find any target detection information embedded in the received wave. The sensor and signal processing constitute the sensing unit in the target tracking system. The second part of a generalised tracking system is a tracker/estimator. The sensing unit provides measurement information to the tracker.

The job of the tracker is to process these measurements and estimate the state of any target. The state of the target can be defined by attributes of the target such as its kinematics, shape and even type of the target. The performance of the estimator is defined by some objective functions with the aim to keep tracking errors as small as possible. The subsystem which performs appropriate control actions based on the tracker output is the third part of a tracking system. An operator observing the target state initiates commands to control the surveillance environment. For example, in civil aviation, the flight controller observes the target tracking results and instructs the targets to behave in a prescribed manner.

The focus of this thesis is to study the second part of a target tracking system, i.e., designing trackers or estimators which process noisy sensor measurements. In 1960, R. E. Kalman introduced modern tracking based on the state space models through his pioneer work on recursive filtering [38].

2.2 Bayesian Tracking Problem Formulation

In this section, we formulate the problem of tracking in a state space framework and provide solutions based on Bayesian statistical inference. We start with a mathematical description of the stochastic filtering problem and then reduce this to the special linear Gaussian case. Consider the following dynamic system and observation equation

$$\begin{aligned} \dot{x}_t &= f(x_t, u_t, \nu_t, t) \\ y_t &= h(x_t, u_t, w_t, t) \end{aligned} \tag{2.1}$$

where x_t represents the state vector of the system with dimension n, y_t represents the measurement vector with dimension p , u_t is the system input, ν_t is the process noise entering the system, w_t is the measurement noise encountered due to sensor imperfections and t is time. The functions f and h are state transition and measurement functions defined as $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ and $h : \mathbb{R}^n \mapsto \mathbb{R}^p$ respectively.

The dynamical system defined above in equation (2.1) contains the state evolution equation and the measurement equation in continuous time. For implementation of so-

lutions we need to look at the discrete time form of above system,

$$\begin{aligned} x_{k+1} &= f(x_k, u_k, v_k, k) \\ y_k &= h(x_k, w_k, k) \end{aligned} \tag{2.2}$$

where v_k and w_k are discrete time process and measurement noises respectively and can be considered as white noise sequences. Also, in most of the tracking problems we consider the state equation without input, i.e., $u_k = 0$.

We adopt a Bayesian approach to the development of an estimator for the system described in equation (2.2) as discussed in [32]. The essence of all Bayesian filtering is to update the prior distribution of the state with new measurement information resulting in the posterior distribution of the state. The posterior distribution is proportional to the product of the likelihood of the current measurement and the prior distribution of the state (Bayes rule). The posterior distribution calculated at time $k-1$ becomes the prior distribution at time k and is propagated forward using the target dynamics. The parameters of the posterior distribution such as mean and covariances tell us the performance of our estimator when compared with true parameters. The basic mathematical structure of the Bayesian solution can be written as below.

$$p(x_k|Y^k) = \frac{p(y_k|x_k)p(x_k|Y^{k-1})}{C} \tag{2.3}$$

where $p(x_k|Y^k)$ is the posterior distribution of state x utilising all the measurements received until time k, $p(y_k|x_k)$ is the likelihood of measurement y_k, $p(x_k|Y^{k-1})$ is the predicted probability distribution function of state x and C is a normalisation constant.

We denote the prior distribution of x at time $k-1$ conditioned on the measurement set Y^{k-1} received upto time $k-1$ as $p(x_{k-1}|Y^{k-1})$. The predicted state density $p(x_k|Y^{k-1})$ can be calculated by the Chapman-Kolmogrove equation

$$p(x_k|Y^{k-1}) = \int_{x_{k-1}} p(x_k|x_{k-1})p(x_{k-1}|Y^{k-1})dx_{k-1} \tag{2.4}$$

where $p(x_k|x_{k-1})$ is state transition density and it depends on the mapping function f.

The normalization constant C in equation (2.3) is calculated using

$$C = p(y_k|Y^{k-1}) = \int_{x_k} p(y_k|x_k)p(x_k|Y^{k-1})dx_k \tag{2.5}$$

where $p(y_k|Y^{k-1})$ is called the predicted measurement distribution. The posterior distribution of state x_k becomes

$$p(x_k|Y^k) = \frac{p(y_k|x_k)}{p(y_k|Y^{k-1})} \int_{x_{k-1}} p(x_k|x_{k-1})p(x_{k-1}|Y^{k-1})dx_{k-1} \tag{2.6}$$

Finding the optimal solution of equation (2.6) is intractable in general. However in certain cases we can develop useful recursive algorithms. The first major assumption is that we consider additive process and measurement noises. The discrete state space model described in equation (2.2) with no input, i.e., $u_k = 0$ can be written as

$$\begin{aligned} x_{k+1} &= f(x_k) + \nu_k \\ y_k &= h(x_k) + w_k \end{aligned} \tag{2.7}$$

In this case, the transition density $p(x_k|x_{k-1})$ and likelihood of measurement y_k can be written as

$$\begin{aligned} p(x_k|x_{k-1}) = p_{\nu_k}(f^{-1}(x_k, x_{k-1}))| \bigtriangledown_{x_k} f^{-1}(x_k, x_{k-1})| = p_{\nu_k}(x_k - f(x_{k-1})) \\ p(y_k|x_k) = p_{w_k}(h^{-1}(y_k, x_k))| \bigtriangledown_{y_k} h^{-1}(y_k, x_k)| = p_{w_k}(y_k - h(x_k)) \end{aligned} \tag{2.8}$$

The posterior density of the additive noise model defined in equation (2.7), after substitution of the transition density and likelihood from equation (2.8) in equation (2.6) respectively, becomes

$$p(x_k|Y^k) = \frac{p_{w_k}(y_k - h(x_k)) \int_{x_{k-1}} p_{\nu_k}(x_k - f(x_{k-1}))p(x_{k-1}|Y^{k-1})dx_{k-1}}{\int_{x_k} p_{w_k}(y_k - h(x_k))p(x_k|Y^{k-1})dx_k} \tag{2.9}$$

The Bayesian equation (2.9) is the foundation for many tracking algorithms because in most practical tracking scenarios the disturbances are additive. Until now, we have not made any assumptions about the distributions of the plant and sensor disturbances

and have not discussed the nature (linearity or non-linearity) of the mapping functions.

The closed form solution of equation (2.9) (also known as Kalman filter) can be derived by assuming Gaussian noise models and linear mappings f and h. The first assumption is that the object dynamics and measurement equations are linear and can be written as:

$$\begin{aligned} x_{k+1} &= F_k x_k + \nu_k \\ y_k &= H_k x_k + w_k \end{aligned} \tag{2.10}$$

where F_k is the state transition matrix and H_k is the measurement output matrix with appropriate dimensions. When these matrices are static (not changing with time), we simply drop the subscript k. The second assumption is that the process and measurement noises ν_k and w_k are white Gaussian sequences with means 0 and covariances Q_k and R_k respectively. We also assume the process noise sequence ν_k is uncorrelated with the measurement noise sequence w_k. The density of the state x is Gaussian at every time and can be fully described by two parameters, i.e., mean and covariance matrices. The closed form solution of the dynamic model defined in equation (2.10) with linear Gaussian assumptions is the well known Kalman Filter (KF) presented in Algorithm 1. At time $k-1$, the prior probability distribution function (pdf) of the state x_k can be written as $p(x_{k-1}|Y^{k-1}) = N(x_{k-1}; H\hat{x}_{k-1|k-1}, P_{k-1|k-1})$.

The Kalman filter is optimal (Minimum Means Square Error and Maximum a Posteriori) under linear Gaussian assumptions and is also the minimum variance estimator. There are various forms of the Kalman filter such as the Covariance filter, the Information filter and the Square Root filter as discussed in [2] , [30] and [36].

2.2.1 Non-Linear Bayesian Filtering

In practical target tracking, linear Gaussian assumptions do not always hold and the Kalman filter results are not optimal and the target tracking problem becomes more complex. However, the general Bayesian approach is still applicable and approximations can be developed.

Algorithm 1 : Recursion Cycle of Kalman Filter

1: Prediction Step

$$\hat{x}_{k|k-1} = F\hat{x}_{k-1|k-1}$$
$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q$$

2: Compute Kalman gain K_k, predicted measurement $\hat{y}_{k|k-1}$ and covariance of innovations

$$K_k = P_{k|k-1}H^T S_k^{-1}$$
$$\hat{y}_{k|k-1} = H\hat{x}_{k|k-1}$$
$$S_k = HP_{k|k-1}H^T + R$$

3: Update Step with measurement y_k

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1})$$
$$P_{k|k} = (I - K_kH)P_{k|k-1}(I - K_kH)^T + K_kRK_k^T$$

4: Resulted posterior pdf $p(x_k|Y^k) = N(x_k; H\hat{x}_{k|k}, P_{k|k})$

Two important approaches for solving non-linear target tracking problems are now discussed. The first is linearisation of the non-linear functions f and h as defined in equation (2.7). This linearisation is usually performed using a Taylor series expansion. After performing the linearisation, the Kalman filter equations are used to update the prior pdf. The extended Kalman filter is an example of a linearisation filter using Taylor series expansions. The second approach is to find the moments of the non-linearly transformed random variables. A further method called the Unscented Kalman filter is also described below.

2.2.2 Extended Kalman Filter

The Extended Kalman Filter (EKF) uses a Taylor series approximation to the non-linear functions $f(.)$ and $h(.)$. The distribution of disturbances and state are still considered Gaussian. The propagation and update steps follows the Kalman filter equations. The details of derivations of EKF can be found in [36], [14] and [2].

Linear Approximation of f and h

In this subsection, we review the process of linearisation of non-linear functions f and h defined in equation (2.7). Full details of the complete derivation of EKF can be found in [14].

First we write the Taylor series expansion of $f(x_{k-1})$ at $\hat{x}_{k-1|k-1}$

$$f(x_{k-1}) = \hat{x}_{k-1|k-1} + \sum_{j=1}^{\infty} \frac{1}{j!} (\otimes_{i=1}^{j}) \nabla_{x^T} f(x)|_{x=\hat{x}_{k-1|k-1}} (\otimes_{i=1}^{j} (x_{k-1} - \hat{x}_{k-1|k-1}))$$

where, for a vector $\mathbf{a} = [a_1, a_2, ..., a_n]$, $\nabla_{\mathbf{a}} = [\partial/\partial a_1, \partial/\partial a_2, ..., \partial/\partial a_n]$ and $\otimes$ denotes the Kronecker product. The classical EKF assumes that the first two terms of the Taylor expansion defined above are sufficient and we can neglect the higher order terms. Therefore $f(x_{k-1})$ can be written as

$$f(x_{k-1}) \approx \hat{f}(x_{k-1}) = \hat{x}_{k-1|k-1} + F_k(x_{k-1} - \hat{x}_{k-1|k-1})$$

where $F_k = \nabla_{x^T} f(x)|_{x=\hat{x}_{k-1|k-1}}$

Similarly, we can find the Taylor series expansion of $h(x_{k-1})$ at $\hat{x}_{k-1|k-1}$ as

$$h(x_k) = \hat{x}_{k|k-1} + \sum_{j=1}^{\infty} \frac{1}{j!} (\otimes_{i=1}^{j}) \nabla_{x^T} h(x)|_{x=\hat{x}_{k|k-1}} (\otimes_{i=1}^{j} (x_{k-1} - \hat{x}_{k|k-1}))$$

Ignoring the higher order terms, we can write $h(x_k)$ as

$$h(x_k) \approx \hat{h}(x_k) = \hat{x}_{k|k-1} + H_k(x_{k-1} - \hat{x}_{k|k-1})$$

where $H_k = \nabla_{x^T} h(x)|_{x=\hat{x}_{k|k-1}}$

The EKF is one of the most popular choices for non-linear tracking problems. However the Gaussian approximations used by the EKF are often a poor approximation resulting in unacceptable performance. In such situations, the EKF gives misleading tracking results and therefore we need to look for other solutions. One cycle of EKF is summarised in Algorithm 2.

Algorithm 2 : Recursion Cycle of Extended Kalman Filter

1: Compute the linearisation of f at $\hat{x}_{k-1|k-1}$:

$$F_k = \nabla_{x^T} f(x)|_{x=\hat{x}_{k-1|k-1}}$$

2: Prediction Step

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1})$$
$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k$$

3: Compute the linearisation of h at $\hat{x}_{k|k-1}$:

$$H_k = \nabla_{x^T} h(x)|_{x=\hat{x}_{k|k-1}}$$

4: Compute Kalman gain K_k, predicted measurement $\hat{y}_{k|k-1}$ and covariance of the innovations

$$K_k = P_{k|k-1} H_k^T S_k^{-1}$$
$$\hat{y}_{k|k-1} = h(\hat{x}_{k|k-1})$$
$$S_k = H_k P_{k|k-1} H_k^T + R_k$$

5: Update Step with measurement y_k

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1})$$
$$P_{k|k} = (I - K_k H_k) P_{k|k-1} (I - K_k H_k)^T + K_k R_k K_k^T$$

6: Resulted posterior pdf $p(x_k|Y^k) = N(x_k; H\hat{x}_{k|k}, P_{k|k})$

2.2.3 Unscented Kalman Filter

Julier and Uhlmann [35] proposed a new filter the Unscented Kalman Filter (UKF) using a novel unscented transformation which overcomes some of the limitations of the EKF. The unscented Kalman filter assumes a Gaussian distribution for the state however the Gaussian distribution is approximated using deterministic sampling techniques. The difference between the Monte Carlo approximation of a distribution and the UKF approximation is in the random and deterministic selection of samples, respectively. These samples are called sigma points in the UKF approximations and a selection procedure was firstly proposed in [35].

Algorithm 3 : Recursion Cycle of Unscented Kalman Filter

1: Set the sigma points $\chi^1_{k-1}, \chi^2_{k-1}, ..., \chi^s_{k-1}$, with associated weights $w^1_{k-1}, w^2_{k-1}, ..., w^s_{k-1}$ to match mean $\hat{x}_{k-1|k-1}$ and covariance $P_{k-1|k-1}$ at time $k-1$

2: Perform transformation of sigma points using true non-linear function of state

$$\chi^i_{k|k-1} = f(\chi^i_{k-1})$$

where $i = 1, 2, ..., s$ is index of all sigma points, $s = 2n+1$ is number of sigma points and n is dimension of state vector.

3: Prediction Step

$$\hat{x}_{k|k-1} = \sum_{i=1}^{s} w^i_{k-1} \chi^i_{k|k-1}$$
$$P_{k|k-1} = \sum_{i=1}^{s} w^i_{k-1} (\chi^i_{k|k-1} - \hat{x}_{k|k-1})(\chi^i_{k|k-1} - \hat{x}_{k|k-1})^T + Q_k$$

4: Update sigma points $\chi^1_k, \chi^2_k, ..., \chi^s_k$, with associated weights $w^1_k, w^2_k, ..., w^s_k$ to match mean $\hat{x}_{k|k-1}$ and covariance $P_{k|k-1}$ at time k

5: Perform transformation of sigma points using true non-linear function h of measurement

$$\varkappa^i_k = h(\chi^i_k)$$

6: Compute Kalman gain K_k, predicted measurement $\hat{y}_{k|k-1}$ and covariance of innovations

$$\hat{y}_{k|k-1} = \sum_{i=1}^{s} w^i_k \varkappa^i_k$$
$$S_k = \sum_{i=1}^{s} w^i_k (\varkappa^i_{k|k-1} - \hat{y}_{k|k-1})(\varkappa^i_{k|k-1} - \hat{y}_{k|k-1})^T + R_k$$
$$K_k = \sum_{i=1}^{s} w^i_k (\chi^i_k - \hat{x}_{k|k-1})(\varkappa^i_{k|k-1} - \hat{y}_{k|k-1})^T$$

7: Update Step with measurement y_k

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k S_k^{-1}(y_k - H\hat{x}_{k|k-1})$$
$$P_{k|k} = P_{k|k-1} - K_k S_k^{-1} K_k^T$$

8: Resulted posterior pdf $p(x_k|Y^k) = N(x_k; \hat{x}_{k|k}, P_{k|k})$

In the unscented Kalman filter, sigma points are chosen carefully and their respective weights are calculated. The propagation of these sigma points uses the non-linear state functions and all sigma points are updated using the traditional Kalman filter update procedure. The unscented Kalman filter is outline in Algorithm 3.

Both the EKF and UKF are non-linear filters incorporating a linearisation procedure for the state transition function and measurement output function. Both filters use the same Gaussian distribution for the state and therefore the standard Kalman filter update procedure can be used to find the parameters of the posterior pdf of the state under an additive noise assumption. However, the Gaussian assumption of the state pdf no longer holds in most of the practical problems whose true state is bimodal or heavily skewed. Therefore, the system becomes non-linear and non-Gaussian and the Kalman type equations no longer hold. In order to solve non-linear and non-Gaussian tracking problems, we adopt procedures for approximating a general pdf. Approximate grid based methods are widely used for approximating a pdf using weighted samples.

2.2.4 Approximate Grid Based Methods

In this subsection, approximate grid based methods are discussed for non-linear and non-Gaussian dynamical systems. The measurement and process disturbances are no longer considered Gaussian distributed. The idea is to deal with real problems by relaxing the assumptions made for the previously discussed tracking algorithms.

In approximate grid based methods, we discretize the state space. Consider the state of the system described in equation (2.7). At time $k-1$ the region is divided into n sub-regions denoted as x_{k-1}^i for $i = 1, 2, 3, ..., n$ with associated weight of w_{k-1}^i, with $\sum_{i=1}^n w_{k-1}^i = 1$. Using approximate grid based method, the pdf of the state at time $k-1$ is approximated by

$$p(x_{k-1}|y^{k-1}) \approx \sum_{i=1}^{n} w_{k-1|k-1}^i \delta(x_{k-1} - x_{k-1}^i) \tag{2.11}$$

where $\delta(.)$ is Dirac delta function and x_{k-1} is the true state of the system at time $k-1$.

Algorithm 4 : Recursion Cycle of a Grid Based Approximate Filter

1: Approximate prior pdf of $p(x_{k-1}|y^{k-1})$

$$p(x_k|y^{k-1}) \approx \sum_{i=1}^{n} w_{k|k-1}^{i} \delta(x_k - x_k^i)$$

2: Select grid points at time k, $x_k^1, x_k^2, ..., x_k^n$

3: Compute the predicted weights $w_{k|k-1}^{i}$ for each i

4: **for** $n = 1$ **to** k **do**

5:

$$w_{k|k-1}^{i} = \sum_{m=1}^{n} w_{k-1|k-1}^{i} p_{v_k}(x_k^i - f(x_{k-1}^m))$$

6: **end for**

7: The predicted pdf $p(x_k|y^{k-1})$ can be approximated by with predicted weights:

$$p(x_k|y^{k-1}) \approx \sum_{i=1}^{n} w_{k|k-1}^{i} \delta(x_k - x_k^i)$$

8: Calculate updated weights $w_{k|k}^{i}$ for each i

9: **for** $i = 1$ to n **do**

10:

$$w_{k|k}^{i} = \sum_{m=1}^{n} w_{k|k-1}^{i} p_{w_k}(y_k - h(x_k^i)) / \sum_{m=1}^{n} w_{k|k-1}^{m} p_{w_k}(y_k - h(x_k^m))$$

11: **end for**

12: The updated pdf $p(x_k|y^k)$ can be approximated by with updated weights:

$$p(x_k|y^k) \approx \sum_{i=1}^{n} w_{k|k}^{i} \delta(x_k - x_k^i)$$

13: State estimate can be computed as

$$\hat{x}_{k|k} = \sum_{i=1}^{n} w_{k|k}^{i} x_k^i \tag{2.12}$$

A tracking filter based on approximate grid methods is also called as Point Mass Filter (PMF) and its derivation using a Bayesian framework is included in [15]. Starting with an approximate distribution of the state $p(x_{k-1}|y^{k-1})$ at time $k-1$, the predicted and updated approximate distributions are

$$p(x_k|y^{k-1}) \approx \sum_{i=1}^{n} w^i_{k|k-1}\delta(x_k - x^i_k) \tag{2.13}$$

$$p(x_k|y^{k}) \approx \sum_{i=1}^{n} w^i_{k|k}\delta(x_k - x^i_k) \tag{2.14}$$

One version of the grid based approximate filter is described in the Algorithm 4. The approximation of the posterior pdf depends on the selection of the grid regions. A dense grid will give better performance however the computation load will increase. The problem of selection of the grid, based on the prior distribution is discussed by Bucy et al in [9]. Computational complexity is another major issue in implementing Algorithm 4. Due to these computational issues, grid based methods have failed to attract practical use.

2.2.5 Sequential Monte Carlo Based Filters

In this subsection, we discuss various flavours of Sequential Monte Carlo (SMC) simulation based filters also known as Particle Filters (PF). As discussed above Kalman filters solve the filtering problem for linear-Gaussian systems optimally. However non-linear systems with Gaussian state distributions are sub-optimally solved using methods such as EKF and UKF following the same KF based framework. In subsection 2.2.4, we discussed the brute force grid based approaches to solve the non-linear and non-Gaussian systems. We considered approximating the continuous state space by n discrete partitions and then using an approximation of the posterior pdf of the state with n partitions. This approximation comes with several drawbacks. The computational complexity is an another major hurdle to real time applications in the object tracking.

In order to solve the filtering problem for non-linear and non-Gaussian systems, researchers have exploited Sequential Monte Carlo simulation based methods. The Monte Carlo approach is a simulation based method to approximate the posterior distribution

(usually referred as a target distribution) by using random samples drawn from another distribution also called the sampling distribution or proposal distribution. The weights of the drawn samples are calculated by the ratio of the target and sampling distributions. This procedure is widely known as Importance Sampling (IS) [58]. In a Bayesian estimation scenario, our objective is to find the posterior pdf by having measurement and prior information. Using Sequential Monte Carlo simulation methods, we approximate the posterior pdf by drawing random samples from some sort of proposal distribution and weights are updated using received measurements.

Pioneer work of introducing Sequential Monte Carlo simulation based methods for solving non-linear and non-Gaussian filtering problems was carried out by Handschin et al [29] in 1969. In [29], the authors proposed a novel approach of finding conditional expectation of non-linear and non-Gaussian systems. In 1970, Handschin [28] proposed a framework for recursive tracking using SMC simulation based methods. However, these methods failed to attract widespread use until their re-birth by Gordon et al in 1993. Monte Carlo approximation methods have received huge attention for target tracking and estimation in recent decades due to their ability to handle non-linearities and their computational practicability. Particle filters are the most famous among the Sequential Monte Carlo methods. The basic structure of particle filters is covered in [12] and [1].

The core idea behind the theory of particle filters is approximating estimates by using a weighted sum of particles (samples) and the weights of the particles are calculated by the ratio of the target and sampling distributions. Since the weight update process depends on the target as well as sampling distributions the choice of the sampling distribution is of critical importance. Some of the basic types of sampling densities also known as importance functions are discussed in [1]. Simandle et al [63] discusses the use of various sampling densities and compares their tracking results. The concept of adaptive importance sampling in a particle filter is discussed in [54], [20] and [17]. The limitations of hardware implementation of particle based tracking algorithms are discussed in [63] and are solved in [66] by taking into account the efficiency of the filter by estimating the parameters of the sampling density recursively. The consistency of Monte Carlo approximation type algorithms is important and needs to be addressed for almost every

algorithm based on this method. Heijden [31] discusses consistency checks for particle type filters but not including consistency checks for adaptive importance sampling. The most recent paper by Marin et al [43] discusses the consistency of adaptive multiple importance sampling methods. In this approach, the number of the drawn samples per scan is fixed but with time varying importance density.

There are several variants of the filters using Sequential Importance Sampling (SIS) techniques. The bootstrap filter is one of the first SIS based filters proposed by Gordon et al [27]. The algorithm of the basic SIS based filter is outlined in Algorithm 5 and a derivation can be found in [15].

The performance of particle filters depends on the choice of proposal density as well as the number of samples chosen and has been discussed extensively in the literature [1]. The derivations of various kinds of particle filters are included in [15]. Discussion of the implementation of particle filters is given in [1] with pseudo-code of several variants of the particle filters.

Algorithm 5 : Recursion Cycle of Bootstrap Filter

1: At time $k-1$, we have x_{k-1}^i, w_{k-1}^i for $i = 1, 2, .., n$
2: **for** $i = 1$ to n **do**
3: Re sampling : Find the mixture index p_i using rule $P[p_i = t] = w_{k-1}^t$
4: Draw samples of process noise v_k^i using its distribution
5: Propagate sample x_{k-1}^i

$$x_k^i = f(x_{k-1}^{p_i}) + v_k^i$$

6: Calculate the update weight

$$q_k^i = p_{w_k}(y_k - h(x_k^i)) \tag{2.15}$$

7: **end for**
8: Update and normalize the weights of all samples

$$w_k^i = w_{k-1}^i q_k^i / \sum_{l=1}^{n} w_{k-1}^l q_k^l$$

9: State estimate can be computed as

$$\hat{x}_{k|k} = \sum_{i=1}^{n} w_k^i x_k^i \tag{2.16}$$

2.3 Bayesian Target Tracking in Clutter

In this section, we discuss target tracking in clutter. So far, we have discussed single target tracking with the following assumptions:

1. Targets are point targets
2. Targets are always detectable
3. Sensors receive only target detections
4. No clutter in the surveillance region

However, in practical target tracking problems, we have to deal with cluttered environments and uncertainty associated with the origin of measurements. Sensors receive a set of the measurements including target measurements and clutter measurements. The question of how these measurements contribute to optimal target tracking is answered by data association algorithms. Each measurement plays a role in the update step of state estimation related to its probability of being a target measurement.

Sensor(s) receives a set of detections and not all of these detections originate from valid targets. Sittler introduced the concept of using a validation gate around the predicted state of the measurement in his pioneer work [64]. Only measurements falling inside a validation gate are used to update a track. The gate is a volume of surveillance space centered around the measurement prediction point. The size of the gate is set so that target measurements are highly likely to be found in the gate. Even after applying the gate, there can be zero or more than one measurement inside the gate. Each measurement falling inside the gate can be one of the following three events. All these events are mutually exclusive for the infinite resolution sensor case.

1. Target detection
2. Clutter detection
3. Measurement is from other target/track

In the literature [15], associating measurements to the above events is widely known as the data association problem, i.e., to differentiate between the target and clutter measurements. Following the pioneer work of Sittler, tracking filters based on data association methods have been discussed extensively in the literature, for example [15], [62], [8]

and [7].

Optimal target tracking algorithms based on data association requires all previous scans of measurement informations however it is not possible to retain and process the full history of scan data due to computational limitations. The algorithms discussed in this chapter use both single scan data and multiple scan data for data associations. These algorithms fall under the category of sub-optimal target trackers for cluttered environments.

The simplest data association methods using a validation gate is the nearest neighbour method. In this method, we choose the nearest measurement to the predicted measurement position and this selected measurement is used to update the track state. A filter based on nearest neighbour data association typically uses the Kalman filter equations and is known as the Nearest Neighbour Filter (NNF) [62]. The version of the NNF for multi-target tracking is called the Global NNF (GNN) and is covered in [8]. In the GNN filter, one measurement is assigned to only one track based on minimising the sum of the distances of the assigned measurements to the associated track. There are several drawbacks of filters using nearest neighbour data association techniques including track loss and poor tracking performance for manoeuvring targets.

In the nearest neighbour filter, we use only one measurement and discard other measurements although they may be in close proximity of the predicted measurement position. It is likely that we may not use the true target measurement for updating the track state and instead use one of the clutter measurements. This idea suggests that every measurement inside the track gate can be a target measurement with certain probability. Bar Shalom et el [4] proposed a data association method utilising all measurements inside the track gate. The Probabilistic Data Association Filter (PDAF) is based on evaluating the probability of all measurement being target measurements. The update step of the PDAF filter consists of updating the track state using all measurements inside the validation gate. The resulting conditional estimates are then combined to give one posterior pdf of the state. The PDA filter outperforms NNF in terms of both track loss and tracking performance. The multi target tracking variant of the PDAF is called the Joint PDAF (JPDAF) and was proposed by Bar Shalom in [62]. The JPDA filter discussed in [62] uses

only current scan measurements for joint data associations. In 1993, Rocker proposed a version of the multiple scan JPDA which not only uses the current scan measurements but also previous scans measurements for joint data associations [60]. Both versions of the JPDA filter are used for multiple target tracking however computational requirements increase exponentially with the increase in the number of measurements and number of the targets. Kirubarajan et al [39] presented an overview of the target tracking algorithms using probabilistic data association methods for single as well multiple target tracking in clutter.

Standard probabilistic data association methods (PDA, JPDA) do not include the uncertainties associated with appearance and disappearance of targets in the surveillance space. The existence of a target in a surveillance space is assumed to be certain. However in the practical target tracking problems, targets can be lost, new targets can appear and false tracks can be initialised with clutter measurements. The idea of including the probability of target existence in the probabilistic data association based target tracking systems was introduced by Musicki et al [46]. The filter including the probability of target existence in its design is called the Integrated PDA (IPDA) filter. False track discrimination can be performed in IPDA by examining the track existence statistics. At each scan, the probability of target existence is updated along with the track filter update. The computational requirements for the IPDA filter are only slightly higher than conventional PDA. Similar to the JPDA [62], the joint IPDA (JIPDA) has been developed for multiple target tracking in clutter [45]. In JIPDA, each track is maintained using its probability of target existence and tracks with very low probability of target existence are terminated. The demanding computational requirements of JIPDA led to the development of computational efficient versions of multiple target tracking filters such as Linear JIPDA (LJIPDA) and Linear multitarget IPDA [15].

All IPDA based algorithms discussed so far use only current scan of data. Musicki et al proposed the Integrated Track splitting (ITS) filter for multiple target tracking [48] , [47]. ITS uses more than one scan of the measurement history for updating the probability of track existence within an IPDA based framework. Finite set based target tracking algorithms are also discussed in literature [15].

2.4 Summary

In this chapter, we briefly overviewed target tracking algorithms. We examined the commonly used linear and non-linear filters. Particle filters are also discussed. We included the basic algorithms for target tracking in cluttered environments. In the following chapters, we will discuss distributed tracking and estimation with a focus on algorithms suitable for bandwidth constrained communication links between sensors and global tracking systems.

Chapter 3

Target Tracking Using Scaled Equivalent Measurements

In this chapter, we consider distributed target tracking under communication constraints. This work comprises the first major contribution of this thesis. Performance of tracking improves when using distributed sensor fusion due to the availability of more information. However, we often face bandwidth restrictions for transmission of measurements of locally tracked states to the fusion center. This chapter describes the information fusion problem where we have such bandwidth constraints. We propose that sending measurement components of the full state equivalent measurements at the sensor node has advantages over sending the full state equivalent measurements or local estimates. Global filtering improves when scaled inverse equivalent measurements are used rather than scaled equivalent measurements. Coder operating characteristic curves for a fixed bandwidth channel, show the optimal scan interval for sending the scaled equivalent measurements to the fusion center. Our simulations also show that the optimal number of bits required to encode scaled equivalent measurements is independent of the transmission frequency. We observe that transmitting scaled information with lower encoding errors and lower frequency is better than transmitting with higher encoding errors and higher frequency when there is a constraint on bandwidth.

3.1 Introduction to Distributed Tracking

TARGET tracking with spatially distributed sensors and with constraints on communication resources such as bandwidth, is widely discussed in the literature. In a traditional spatially distributed sensor network, sensors collect target measurements in the surveillance area and transmit them to the fusion center where tracking is performed. The fusion center fuses the received measurements and updates the track of the target being followed. It could be deduced from information theory that more information leads

to better tracking. However sending the measurements to the fusion sensor can be expensive in terms of bandwidth usage. Under constraints on bandwidth, instead of sending whole measurements, we look for parts of measurements which carry useful information and need less bandwidth to transmit.

When the primary objective is tracking performance in the mean square error sense, the optimal architecture for tracking targets using spatially distributed sensors is a centralized architecture, where each node sends its measurements to the fusion center [8]. However, the centralized architecture comes with several drawbacks, consisting of high bandwidth requirements, high vulnerability to attack, delay in transmission and reception, information received out of order etc [40].

Distributed or decentralized tracking overcomes some problems of the centralized architecture. In distributed tracking each local sensor has the capability to perform local tracking using its own measurement data and sends local state estimates to the fusion center. Nonetheless, the problem becomes complex when received estimates from various sensors are correlated with each other [41].

Transmitting local estimates in a decentralized architecture requires less bandwidth when compared to a centralized structure. However, bandwidth requirements can be further reduced by transmitting equivalent measurements and/or equivalent innovations as discussed in [40] and [44]. Transmitting equivalent measurements effectively saves bandwidth and proves to be a better approach as discussed in [6],[24], [21] and [22]. In a decentralized architecture, degradations of tracking performance have been discussed in [72] and [73]. In a centralized architecture, problems of bandwidth allocation are discussed in [65] and [61].

Although sending equivalent measurements instead of sensor measurements provides a significant decrease in bandwidth utilization, sending equivalent innovations further saves bandwidth. However equivalent innovations transmission leads to random walk phenomenon as discussed in [44] due to accumulation of encoding errors at the fusion center. Transmitting equivalent measurements has advantages in the sense that we send only useful information extracted from measurements. However all components of the equivalent measurement (full state equivalent measurement) do not carry equally

useful information. Therefore in our work, we propose to send only measurement components of the equivalent measurements termed scaled equivalent measurements, which further saves communication bandwidth.

The choice of filter at the fusion center is independent of the local filter. Global filtering at the fusion center performs better when using inverse scaled equivalent measurements as compared to scaled equivalent measurements. Therefore we use an information filter update form at the fusion center using inverse scaled equivalent measurements.

We save bandwidth by reducing transmission frequency. We also propose that instead of transmitting information every scan, we transmit after N scan intervals and observe only a small performance degradation. Since full state equivalent measurement needs more bits per scan to send as compared to scaled equivalent measurements, we can send scaled equivalent measurements more frequently using the same bandwidth. We observe that both full state and scaled equivalent measurements produce near optimal performance when transmitting every scan with no bandwidth constraints. Our simulations demonstrate that under bandwidth constraints sending scaled equivalent measurement more frequently is better, when compared to sending full state equivalent measurements less frequently .

The challenge in tracking using sensor networks is to utilize communication bandwidth in an efficient way to keep tracking Root Mean Square Errors (RMSE) as small as possible. However, we face larger errors than the optimal due to the loss of information in encoding, transmission, and channel noise. We propose another algorithm to utilize bandwidth efficiently. We incorporate a strategy of not transmitting the scaled equivalent measurements on each scan but at some scan interval. Doing so, performance degradation is observed due to loss of information but it is small. This problem becomes more interesting, when we are allocated fixed bandwidth channels to be used for each target to be tracked for a certain time duration. We find that the performance of tracking depends on the minimum number of bits used to encode data [40]. Therefore for a fixed bandwidth budget, we allocate bits per scan based on the frequency of transmission. We observe the optimal number of bits required to encode scaled equivalent measurements is independent of the frequency of transmission. We use coder operating characteristics

to find the optimal transmission scan interval for scaled equivalent measurements under fixed bandwidth.

3.2 Problem Statement

A two dimensional surveillance space is considered. A single sensor scans the surveillance space of area A_S. The sensor under consideration performs local data processing, i.e., local tracking. The sensor platform sends track data to the fusion center where global tracks are maintained. We consider bandwidth constrained communication channel(s) between the sensor and the fusion center. Figure 3.1 shows the structure of the system under consideration. In the literature, this architecture falls under distributed or decen-

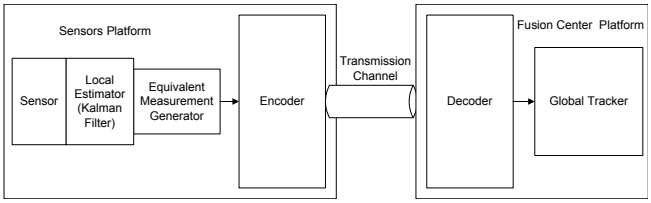

Figure 3.1: Overview of the decentralized tracking architecture

tralised tracking [41], and is useful under bandwidth constrained situations. Distributed tracking provides flexibility over the choice between sending local estimates, equivalent measurements or equivalent innovations.

We consider a track following the target trajectory described by

$$x_{k+1} = Fx_k + v_k \tag{3.1}$$

where k represents sensor scan time, F is the target state transition matrix, and process noise v_k is a zero mean white Gaussian noise sequence with covariance matrix Q. The sensor measurement model is

$$y_k = Hx_k + \omega_k \tag{3.2}$$

where H denotes the measurement matrix and measurement noise ω_k is a zero mean white Gaussian noise sequence with covariance matrix R uncorrelated with process noise

ν_k.

The traditional Kalman filter maintains local tracks at sensor nodes using sensor measurements y_k. For simplicity, we consider a single sensor and a single fusion center connected via a bandwidth constrained communication channel. At each scan time k, the sensor provides measurement y_k and the local Kalman filter updates the track state according to the standard Kalman filter recursion cycle given in equation (3.3).

$$\begin{aligned} \hat{x}_{k|k-1} &= F\hat{x}_{k-1|k-1} \\ P_{k|k-1} &= FP_{k-1|k-1}F^T + Q \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1}) \\ P_{k|k} &= (I - K_kH)P_{k|k-1}(I - K_kH)^T + K_kRK_k^T \end{aligned} \tag{3.3}$$

where $\hat{x}_{k|k}$ denotes the state estimate, $P_{k|k}$ the state error covariance matrix, I is the identity matrix, $(.)^T$ is the transpose of a matrix and K_k is the Kalman gain, calculated by

$$K_k = P_{k|k-1}H^T[HP_{k|k-1}H^T + R]^{-1}$$

We calculate the equivalent measurements by employing an inverse filter at the output of the Kalman filter and then convert this to scaled equivalent measurements at the local sensor. We use expressions for equivalent measurements derived in [50] and modify these to N-step equivalent measurements $Z_{\varepsilon,N}$ and covariance matrix $R_{\varepsilon,N}$, where N is a positive non-zero integer. The frequency of transmission is $\frac{1}{N}$. The local sensor transmits N-step scaled equivalent measurements $Z^s_{\varepsilon,N}$ and covariance matrix $R^s_{\varepsilon,N}$ to the fusion center after compressing the information with a finite number of bits.

The fusion center receives encoded strings of $Z^s_{\varepsilon,N}$ and $R^s_{\varepsilon,N}$ and decodes these into $\hat{Z}^s_{\varepsilon,N}$ and $\hat{R}^s_{\varepsilon,N}$. After decoding , we convert $\hat{Z}^s_{\varepsilon,N}$ and $\hat{R}^s_{\varepsilon,N}$ into the full state inverse equivalent measurements, i.e., $\hat{y}_{\varepsilon,N}$ and covariance Matrix $\hat{R}^{-1}_{\varepsilon,N}$. The fusion center uses these decoded inverse equivalent measurements to update the global information filter. The next section explains fusion with the full state and scaled equivalent measurements.

3.3 Fusion with Equivalent and Scaled Equivalent Measurements

If we consider a Gaussian probability distribution function with error covariance matrix Γ, then the uncertainty volume is proportional to the square root of the determinant of the error covariance matrix, i.e., $\sqrt{|\Gamma|}$ as described in [44]. The number of bits required to encode information with covariance matrix Γ is directly proportional to $\sqrt{|\Gamma|}$. Optimal fusion uses a centralized architecture in which sensors transmit whole measurements to the fusion center [8]. However due to bandwidth constraints, we transmit equivalent measurements and scaled equivalent measurements. Therefore the global filter type does not need to be the same as the local filter.

In this section, we further discuss fusion with full state and scaled equivalent measurements. As already stated above, in a centralized configuration, transmission of sensor measurements is expensive in terms of bandwidth. Furthermore in a distributed architecture, transmitting local estimates $\hat{x}_{k|k}$ and $P_{k|k}$ not only requires a high number of encoding bits but is also vulnerable to various forms of data loss [40]. On the other hand, sending equivalent innovations requires relatively few bits as compared to the full state equivalent measurements but errors accumulates at the fusion center as discussed in [44]. Therefore, we choose equivalent measurement information as a 'good choice' candidate for track state transmission.

We calculate the N-step equivalent measurements $Z_{\varepsilon,N}$ and covariance matrix $R_{\varepsilon,N}$ as follows [50],

$$\begin{aligned} R_{\varepsilon,N}^{-1} &= P_{k|k}^{-1} - P_{k|k-N}^{-1} \\ y_{\varepsilon,N} &= P_{k|k}^{-1}\hat{x}_{k|k} - P_{k|k-N}^{-1}\hat{x}_{k|k-N} \\ R_{\varepsilon,N} &= (R_{\varepsilon,N}^{-1})^{-1} \\ Z_{\varepsilon,N} &= R_{\varepsilon,N} y_{\varepsilon,N} \end{aligned} \tag{3.4}$$

However, to reduce bandwidth utilization, we transmit only position components of the equivalent measurement and achieve comparable tracking performance to that achieved when full state equivalent measurement are transmitted. This requires less bandwidth

and less transmission time. Instead of transmitting $Z_{\varepsilon,N}$ and covariance matrix $R_{\varepsilon,N}$, we scale this information with the H matrix and transmit only those components of $Z_{\varepsilon,N}$ and $R_{\varepsilon,N}$ which correspond to the position components of the state. We calculate scaled $Z^s_{\varepsilon,N}$ and covariance matrix $R^s_{\varepsilon,N}$ as

$$\begin{aligned} R^{-1s}_{\varepsilon,N} &= HR^{-1}_{\varepsilon,N}H^T \\ y^s_{\varepsilon,N} &= Hy_{\varepsilon,N} \\ R^s_{\varepsilon,N} &= (R^{-1s}_{\varepsilon,N})^{-1} \\ Z^s_{\varepsilon,N} &= R^s_{\varepsilon,N}y^s_{\varepsilon,N} \end{aligned} \tag{3.5}$$

The sensor node encodes both the full state equivalent measurements and the scaled equivalent measurements and transmits to the fusion center. The fusion center receives, decodes and uses information as measurements to update the global tracker. We assume that the transmission network is noiseless, reliable and there is no time delay for transmission. We also assume that no out of order transmission occurs. We use the traditional information filter [2] as a global tracker and update this filter using the full state decoded equivalent measurements as follows.

$$\begin{aligned} P^{-1}_{k|k} &= P^{-1}_{k|k-N} + (\hat{R}_{\varepsilon,N})^{-1} \\ P^{-1}_{k|k}\hat{x}_{k|k} &= P^{-1}_{k|k-N}\hat{x}_{k|k-N} + \hat{R}^{-1}_{\varepsilon,N}\hat{Z}_{\varepsilon,N} \end{aligned} \tag{3.6}$$

However, for the case of scaled equivalent measurements, the update-step of the Information filter is modified as below.

$$\hat{R}^{-1s}_{\varepsilon,N} = (\hat{R}^s_{\varepsilon,N})^{-1}\hat{y}^s_{\varepsilon,N} = \hat{R}^{-1s}_{\varepsilon,N}\hat{Z}^s_{\varepsilon,N} \tag{3.7}$$

$$P^{-1}_{k|k} = P^{-1}_{k|k-N} + H^T\hat{R}^{-1s}_{\varepsilon,N}H \tag{3.8}$$

$$P^{-1}_{k|k}\hat{x}_{k|k} = P^{-1}_{k|k-N}\hat{x}_{k|k-N} + H^T\hat{y}^s_{\varepsilon,N} \tag{3.9}$$

Fusion with the scaled equivalent measurements and full state equivalent measurements achieves the same tracking performance as that observed with no bandwidth con-

straints. We convert the received scaled equivalent measurements into the inverse full state equivalent measurements and use the above information filter. We observe performance degradation when using received scaled equivalent measurements in the traditional Kalman filter as the global filter. Figure 3.2 shows near optimal performance of fusion with the equivalent measurements and scaled equivalent measurements when transmitting on every scan.

3.4 Is it Better to Use Equivalent or Scaled Equivalent Measurements

A fusion with both equivalent measurements and scaled equivalent measurements is near optimal (as we discussed in previous section) under no bandwidth constraint. However, the transmission of the equivalent measurements is still more bandwidth expensive when compared with the scaled equivalent measurements. So, under bandwidth constraints a fusion with the scaled equivalent measurements could be a better choice.

For example using a four dimensional state vector, transmission of $Z_{\varepsilon,N}$ and $R_{\varepsilon,N}$ requires fourteen integers, however transmission of $Z^s_{\varepsilon,N}$ and $R^s_{\varepsilon,N}$ requires only five integers. Therefore we roughly observe that the full state equivalent information requires three times bandwidth as compared to its scaled version. We can send the scaled equivalent measurements every scan or the full state every third scan utilizing the same bandwidth. Simulation results of transmitting full state and scaled equivalent measurements at every scan is given in Fig. 3.2. Figure 3.3 shows that transmitting $Z^s_{\varepsilon,N}$ and $R^s_{\varepsilon,N}$ every scan, i.e., with $N = 1$ is better than transmitting $Z_{\varepsilon,N}$ and $R_{\varepsilon,N}$ every third scan, i.e., with $N = 3$. Similarly, sending scaled equivalent measurements with $N = 3$ outperforms sending the full state equivalent measurements with $N = 9$ as shown in Fig. 3.4. For a fixed bandwidth budget, the transmission frequency of the scaled equivalent measurements is higher than the full state equivalent measurements.

Furthermore, bandwidth can be reduced by transmitting the scaled equivalent measurements after N scan intervals. We also discuss the performance comparison of the algorithms transmitting N-step scaled equivalent measurements with a different number

of encoding bits. Coder operating characteristics give us the optimal value of N for a given bandwidth budget. In distributed tracking, there exists a minimum data rate to achieve tracking performance at a certain level of accuracy [40]. Also we observe that this minimum data rate is independent of the frequency of transmission.

We attempt to achieve the best tracking performance with the minimum number of encoding bits for $Z^s_{\varepsilon,N}$ and $R^s_{\varepsilon,N}$. Consider a case when we are allocated a bandwidth budget W bits for C seconds of tracking. We define B as the number of encoding bits per scan proportional to the value of N as

$$B = N \times \frac{W}{C}$$

We achieve near optimal tracking performance by sending the scaled equivalent measurements every scan, i.e., with $N = 1$ and under no bandwidth constraint. The tracking performance degrades as N increases in spite of using infinite bandwidth. However under bandwidth constraints, performance depends on the number of encoding bits. We try to find the optimal value of N for given W bits of bandwidth for C seconds of tracking time.

More frequent transmission with lower encoding bits produces poor tracking performance mainly due to higher encoding errors when compared to the case of sending less often with higher number of encoding bits. Therefore for each value of N, there exists a lower limit of bits to achieve acceptable performance and also an upper limit. We also observe that the optimal value of the number of encoding bits does not depend on the frequency of transmission. Simulation results support these claims. This scheme is useful when there is a trade-off between tracking performance and bandwidth. Instead of sending more often with large encoding errors, it is better to send less often with lower encoding errors. This scheme can be easy extended for multi sensor fusion and expected to perform extremely well under bandwidth constraints.

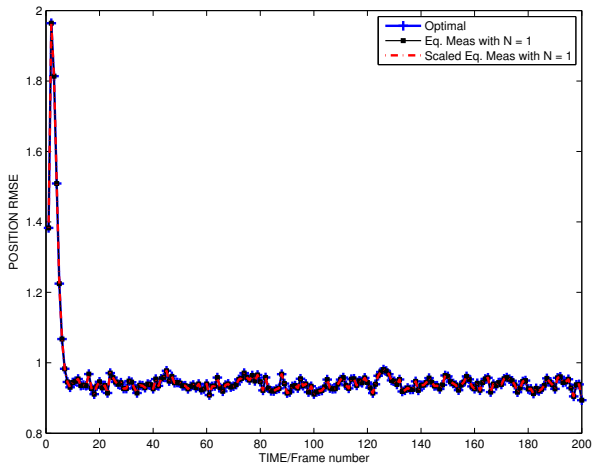

Figure 3.2: Fusion with the equivalent and scaled equivalent measurements when transmitting every scan

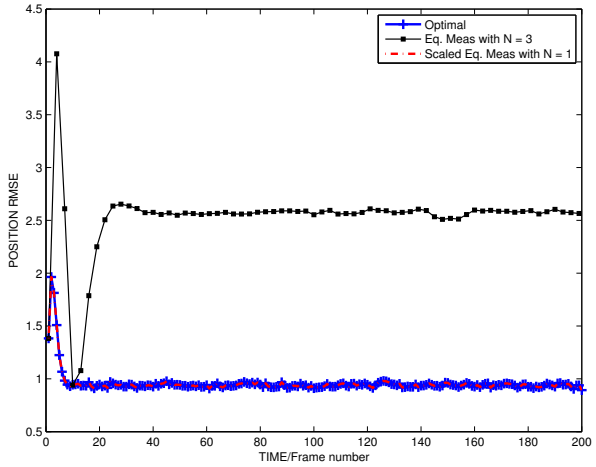

Figure 3.3: Fusion with the equivalent and scaled equivalent measurements

3.5 Simulation Results

Each experiment discussed in this section assumes an identical simulation environment. One target is moving in a straight line in a two dimensional surveillance space. The target is moving with uniform velocity of 5 m/s parallel to the Cartesian x axis. This surveillance space is observed with one sensor with scan interval of $T = 1s$. Sensor measurement noise is assumed to be Gaussian, white and independent of process noise, with covariance matrix R equals to $25I_2\ m^2$ where I_2 represents two dimensional identity matrix. The sensor node is connected to the fusion center via a digital communication

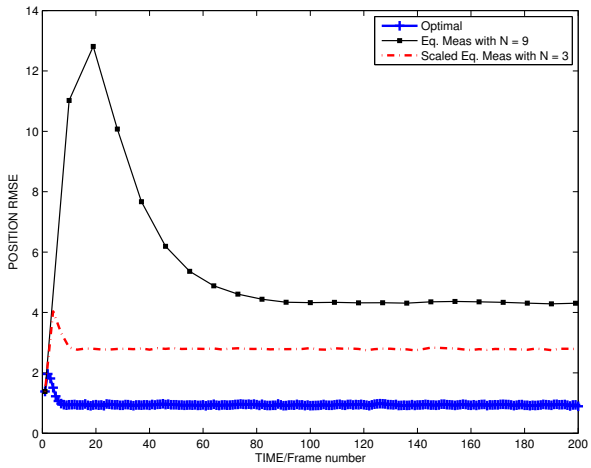

Figure 3.4: Fusion with the equivalent and scaled equivalent measurements using different N

channel with an assumption that the network is ideal with no transmission errors and no out of sequence records.

We use the traditional Kalman filter as a local tracker while the Information filter as a global tracker [2]. Both estimators use Gaussian, white plant noise ν_k with covariance matrix Q defined as defined below

$$Q = 2 \times \begin{bmatrix} \frac{1}{3}T^3 & \frac{1}{2}T^2 & 0 & 0 \\ \frac{1}{2}T^2 & T & 0 & 0 \\ 0 & 0 & \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ 0 & 0 & \frac{1}{2}T^2 & T \end{bmatrix} \tag{3.10}$$

where T is the scan interval time between the previous update step and the current simulated time.

In this section each simulation consists of 1000 simulation runs and each run simulates 200s of the surveillance time. Initialization of the local tracker is carried using one point initialization [49], while initialization of the global tracker is carried out in the same way by transmitting the initialized local track to the fusion center at the start of the tracking time. By doing so, we ensure the same initial conditions for both local and global trackers. This initialization of the local and the global trackers is independent for each simulation run. The tracking performance of the local tracker is considered optimal in the mean

square sense and defined as below.

$$\text{RMSE} = \sqrt{\frac{1}{\text{nRuns}} \times \sum_{i=1}^{\text{nRuns}} (\hat{x}_i - x)^2} \tag{3.11}$$

where RMSE is the position root mean square error, nRuns is the number of runs for simulation, $\hat{x}$ is the estimated position and x is the true position.

We present two sets of simulation experiments. The first set compares the performance of transmitting the full state equivalent measurements and the scaled equivalent measurements. The second set helps us in finding the optimal value of scan interval N for a given bandwidth budget by transmitting the scaled equivalent measurements.

Full State and Scaled Equivalent Measurements

In this subsection, simulation results show that sending the scaled equivalent measurement with a higher frequency has advantages over sending the full state equivalent measurement with a lower frequency using the same bandwidth. The results provided in this section assume that encoded data is received with no errors . We compare results on the basis of RMSE performance criteria defined in equation (3.11). As stated earlier, we require four digits for $Z_{\varepsilon,N}$ and ten digits for $R_{\varepsilon,N}$. In total, we need fourteen digits to send the equivalent measurement per scan. On the other hand, We require two digits for $Z^s_{\varepsilon,N}$ and three digits for $R^s_{\varepsilon,N}$ and total of five digits to send the scaled equivalent measurement per scan. Based on the above statistics, the scaled equivalent measurement requires three times less bandwidth as compared to the transmission of the full state equivalent measurement. This calculation helps us to send the scaled equivalent measurements with three times higher frequency than that of the full state equivalent measurements using the same bandwidth.

We observe that the tracking performance of transmitting the full state equivalent measurements and the scaled equivalent measurements on every scan (i.e., with $N = 1$) is equal to the local tracker performance as shown in Fig. 3.2. Figure 3.2 shows all three performance curves for the local filter, the full state equivalent measurements and the scaled equivalent measurements. These curves overlap. Figure 3.2 also shows that send-

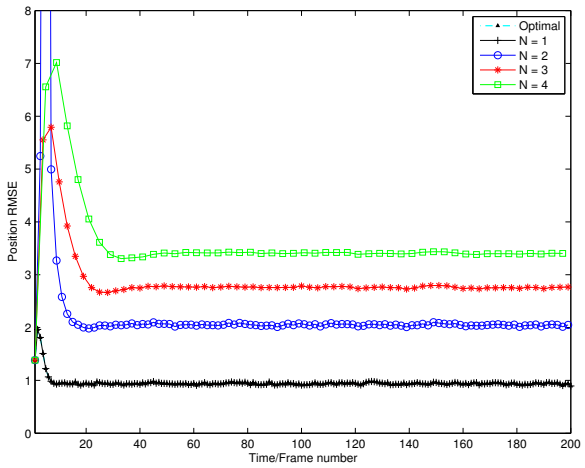

Figure 3.5: Fusion with the N-step scaled equivalent measurements

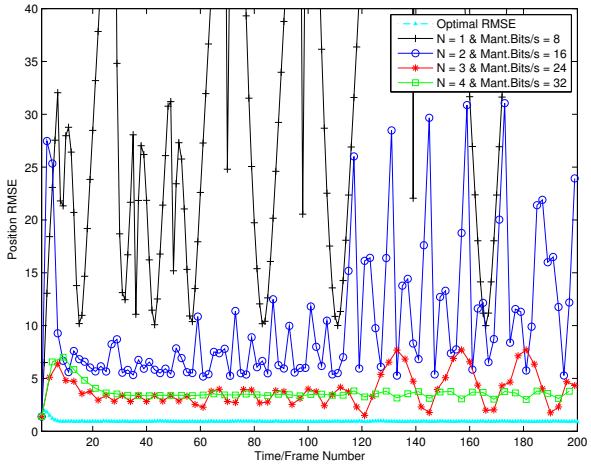

Figure 3.6: Fusion with the scaled equivalent measurements and $W = 1600$ bits

ing the scaled equivalent information carries the same measurement information and uses less bandwidth when compared with the full state equivalent measurements. Figure 3.3 compares the performance of sending the full state equivalent measurements on every third scan and the scaled equivalent measurements on every scan. Also sending scaled measurements on every third scan outperforms sending full state on every ninth scan as shown in Fig. 3.4. These results lead us to the conclusion that we reduce bandwidth usage by transmitting the scaled equivalent measurements and further reduce it by incorporating a strategy of controlling transmission frequency at the cost of minor performance loss.

Sending Scaled Equivalent Measurements and Fixed Bandwidth

Decreasing the frequency of transmission degrades tracking performance, however the bandwidth requirement also reduces. Therefore for fixed bandwidth, transmitting scaled equivalent measurements, which needs less bandwidth, performs better because we can send more frequently. In this subsection, simulation results show the optimal value of N which provides a reasonable trade-off between bandwidth utilization and tracking performance under bandwidth constraints.

Figure 3.5 compares tracking performance using the scaled equivalent measurements with different values of N without encoding. We observe that as N increases, performance degrades to a certain degree. This is because we loose some information when sending with higher scan interval.

In the next set of simulation experiments, we transmit the encoded scaled equivalent measurements, decode them at the fusion center and use them for the global tracker update. We compare algorithms that send scaled equivalent measurements with $N = 1$, $N = 2$, $N = 3$ and $N = 4$ using B encoding bits calculated by equation (3.4), where N denotes transmission scan interval and B is the number of bits per scan. We encode all the above four cases with the same bandwidth budget W for C seconds of tracking, hence each algorithm uses $N \times \frac{W}{C}$ bits to encode scaled equivalent measurements. We use a floating point encoding scheme using two bits for each exponent of $Z^s_{\varepsilon,N}$ and $R^s_{\varepsilon,N}$, while the mantissa bits for $Z^s_{\varepsilon,N}$ and $R^s_{\varepsilon,N}$ are allocated based on available bandwidth W. The maximum value used to encode $Z^s_{\varepsilon,N}$ and $R^s_{\varepsilon,N}$ is 1100 and 25 respectively. We perform several experiments with a variable bandwidth budget W and find the best transmit interval N by looking at the coder operating characteristic curve for each W.

We use a bandwidth budget as shown in Table 3.1 for the above four experiments. Figure 3.6 compares the tracking performance of all four algorithms. For this experiment we use a total of $W = 1600$ mantissa bits for $C = 200$ seconds and distributed according to Table 3.1. The performance curve of the algorithm using $N = 1$ diverges due to the very small number of bits used and it improves with an increasing number of bits as shown in the subsequent figures. Similarly Figures 3.7, 3.8 and 3.9 show the tracking performance with $W = 2400$, $W = 3200$ and $W = 6400$ bits respectively. It is observed that for a lower

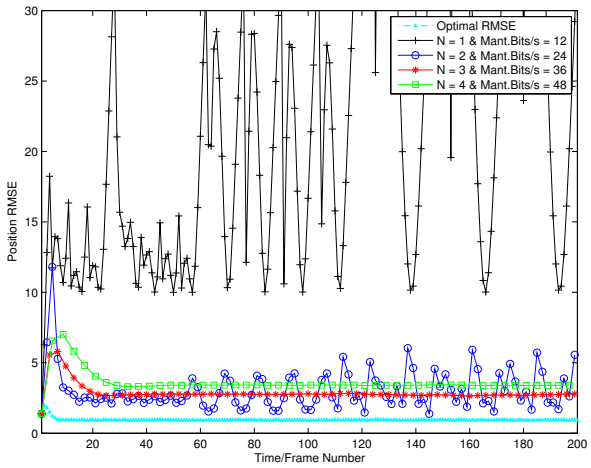

Figure 3.7: Fusion with the scaled equivalent measurements and $W = 2400$ bits

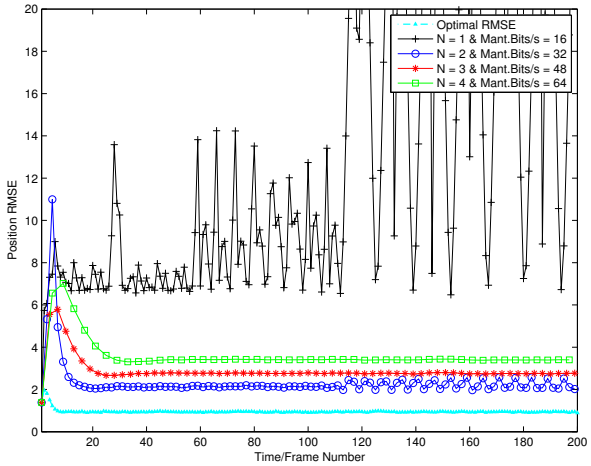

Figure 3.8: Fusion with the scaled equivalent measurements and $W = 3200$ bits

bandwidth budget, sending with a greater N step is better. At this point we see that a trade-off exists between tracking performance and bandwidth constraint. For each value of N, there exists a lower bandwidth limit for acceptable performance. After analysing our results we find the optimal number of encoding bits is between $30 \sim 35$ per scan for achieving reasonable performance for each N. Also this optimal number is independent to the frequency of transmission. Sending less information but with fewer encoded errors is better than sending more information with higher encoding errors.

We also present coder operating characteristics curve which tells us that with an increasing bandwidth budget, performance of the algorithm with lower N increases as

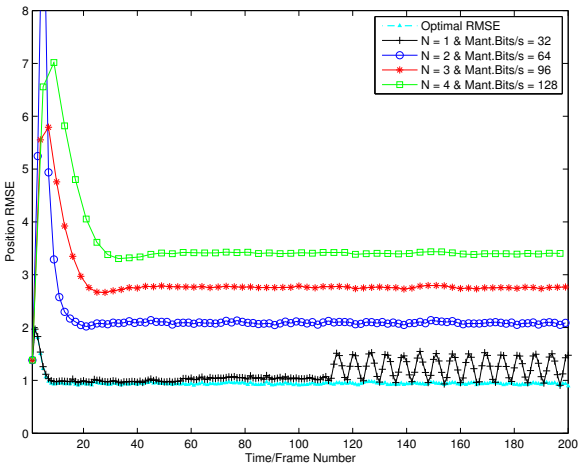

Figure 3.9: Fusion with the scaled equivalent measurements and $W = 6400$ bits

Table 3.1: The value of N and number of bits used for encoding transmitted data under given bandwidth budget

N	$B_{W=1600}$	$B_{W=2400}$	$B_{W=3200}$	$B_{W=6400}$
1	8	12	16	32
2	16	24	32	64
3	24	36	48	96
4	32	48	64	128

shown in Fig. 3.10. This is because encoding errors are reduced. Figure 3.10 gives the optimal value of N to be used for a given bandwidth budget W for $C = 200$ seconds tracking of the target. For example if we have $W = 3200$ bits then the best algorithm uses $N = 2$. We calculate the 'sweet spot' for the trade-off between bandwidth and tracking performance and select the optimal value of N under a given bandwidth budget.

3.6 Summary

In this chapter, we presented a scheme for track state transmission in a distributed tracking architecture. This scheme proved to be helpful in transmitting information with low data rates. We showed that sending the scaled equivalent measurements required a lower number of bits as compared to the full state equivalent measurement transmission while

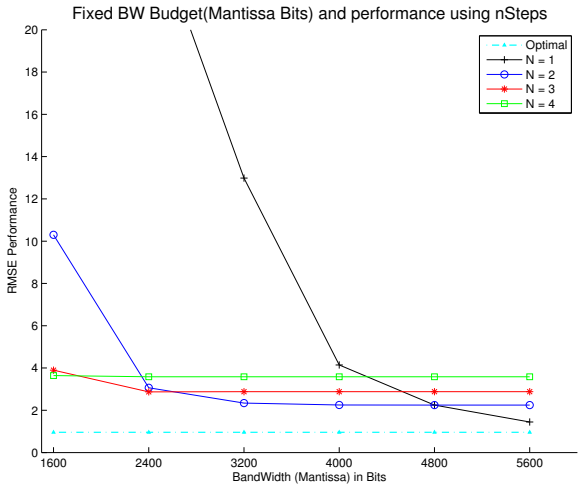

Figure 3.10: Coder operating characteristics under bandwidth constraints

maintaining similar tracking performance. We can send the scaled equivalent measurement with higher transmission frequency rather than send the full state equivalent measurements and we can achieve better performance. We also observed that the global filter performed better when updating is performed using the inverse reduced equivalent measurements.

Under bandwidth constrained situations, it is better to send less often with a higher number of the encoding bits rather than sending more often with a lower number of the encoding bits. We proposed sending information with an N-step scan interval rather than on every scan. By doing so, we distributed the bandwidth to encode the scaled equivalent measurements based on the frequency of transmission. The algorithm using a lower N-step used a lower number of encoding bits per scan and the algorithm using a higher N-step used a higher number of encoding bits per scan. We also discussed the trade-off between tracking performance and bandwidth. We determined the optimal value of N for an allocated bandwidth using the coder operating characteristics. Our simulations also showed that $B = 30 \sim 35$ is the optimal number of bits per scan and is independent of the frequency of the transmission.

Future work is to devise the schemes for the adaptive selection of N based on the information contribution of the equivalent measurements.

Chapter 4

Target Tracking in Clutter Using Compressed Measurement

In distributed tracking, communication bandwidth is one of the most expensive resources when we are required to send measurements to multiple locations for processing. Furthermore, bandwidth requirement increases when tracking in clutter is considered due to transmission of target as well as clutter measurements. This chapter describes tracking in clutter under bandwidth constraints. The main idea is that instead of sending all target and clutter measurements we combine them into a weighted sum and transmit the resultant measurement. A novel Bayesian filter is proposed utilising the received measurement information. We observe some small loss of performance as compared to the local tracker.

4.1 Introduction

TARGET tracking without clutter , with spatially distributed sensors and with constraints on communication resources such as bandwidth, is discussed in the previous chapter. In this chapter, we study distributed target tracking in clutter under bandwidth constraints. In a traditional spatially distributed sensor network, sensors collect target as wells as clutter measurements in the surveillance area and transmit them to the fusion center for global target tracking. The fusion center receives all target and clutter measurements and obtains the global state of the track. The clutter measurements are false alarms, i.e., we perceive detections from non-target objects. Instead of only one target detection, we receive multiple target detections at each scan. Under a constraint on bandwidth, instead of sending target and clutter measurements, we propose combining of target and clutter measurements into a weighted sum and transmit this resultant

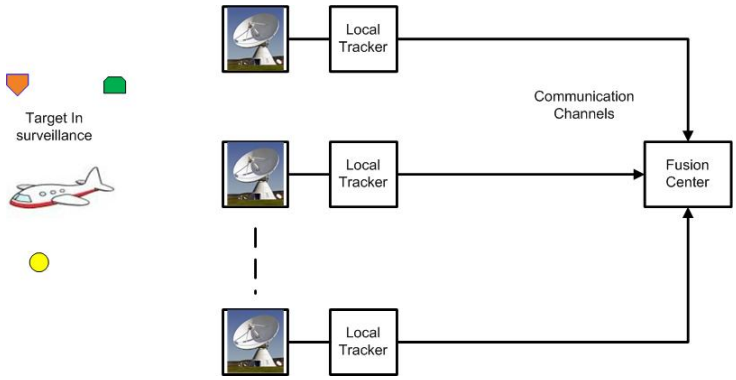

Figure 4.1: Multisensor distributed tracking in clutter

measurement information to the fusion center.

The performance of distributed multitarget tracking in clutter can be studied in terms of the quality of tracks and false track discrimination. In this study, we consider tracking performance as the primary objective and false track discrimination as a secondary objective.

The typical distributed tracking system in clutter is shown in Fig. 4.1. The colored objects in surveillance represent clutter/false targets. We assume the radar is able to detect these objects as targets with probability of target existence is equals to 1. We combine received measurements at the local node using a weighted sum of target and clutter measurements. The weights are calculated based on the likelihood of each measurement. This resultant compressed measurement contains information on clutter and target measurements. We transmit this compressed measurement to the Fusion Center (FC). We propose a novel Bayesian filter which extracts useful information from the compressed measurement at the FC. The proposed filter requires the computation of intractable integrals which are solved with help of Monte Carlo simulations.

The major contribution of this chapter consists of the derivation of the novel Bayesian filter which operates on a weighted sum of target and clutter measurements. By adopting this approach we save communication bandwidth and are able to effectively send both target and clutter measurement information. We show that the loss in tracking performance is almost negligible for our proposed filter when compared to optimal local Probabilistic Data Association Filter (PDAF). The proposed filter uses the compressed

measurement to update track state and hence we call it the Compressed Filter (CF).

4.2 Target Tracking Using Compressed Measurement

We consider distributed tracking where we use local processing of measurements to maintain local tracks and a global track is maintained by receiving measurement information from the local processing units. We consider tracking of a single target moving in a linear trajectory in a one dimension surveillance space in a cluttered environment as follows:

$$x_{k+1} = Fx_k + \nu_k \tag{4.1}$$

where k represents sensor scan time, F is the target state transition matrix and the process noise ν_k is a zero mean white Gaussian noise sequence with covariance matrix Q. The sensor measurement model is

$$y_k = Hx_k + \omega_k \tag{4.2}$$

where H denotes the measurement matrix and the measurement noise ω_k is a zero mean white Gaussian noise sequence with covariance matrix R. The measurement noise is uncorrelated with process noise. The target under observation is detectable with probability of detection equal to one. The clutter is uniformly distributed in the surveillance space. Due to clutter, we receive M_k measurements with at least one target measurement at each scan. The number of clutter measurements per scan follows a Poisson distribution with clutter measurement density ρ. We use the classical probabilistic data association filter proposed in [5] to maintain the local track. The set of validated measurements using a gating procedure is

$$Y_k = \{y_{k,i}\}_{i=1}^{m_k} \tag{4.3}$$

Rather than use a bandwidth to transmit all of the above measurements to the global processing unit, we combine these measurements into one measurement.

$$\breve{y}_k = \sum_{i}^{m_k} \alpha_i y_{k,i} \tag{4.4}$$

where α_i are scalar weights whose values are calculated based on the likelihood of each measurement falling inside the gate. We transmit $\breve{y}_k$ along with the associated weights α_i on every scan to the FC.

4.3 The Compressed Filter

In this section, we derive a filter which uses the combined measurement $\breve{y}_k$ defined in equation (4.4) as a measurement information. We use standard Bayesian procedures to derive this filter as outlined in [14] and call it the Compressed Filter.

The conditional density of x_k given measurement set $\breve{Y}^k = \breve{y}_k, \breve{y}_{k-1}, ..., \breve{y}_1$ (in our case combined measurement) up to time k is denoted as $p(x_k|\breve{Y}^k)$ and defined using Bayesian rule,

$$p(x_k|\breve{Y}^k) = \frac{p(\breve{y}_k|x_k)p(x_k|\breve{Y}^{k-1})}{p(\breve{y}_k|\breve{Y}^{k-1})} \tag{4.5}$$

Since $p(\breve{y}_k|\breve{Y}^{k-1})$ is a constant, we have

$$p(x_k|\breve{Y}^k) \propto p(\breve{y}_k|x_k)p(x_k|\breve{Y}^{k-1}) \tag{4.6}$$

From equation (4.4), $\breve{y}_k$ is a combination of m_k measurements, hence the conditional density can be written to show this explicitly as

$$p(x_k|\breve{Y}^k, m^k) \propto p(\breve{y}_k, m_k|x_k)p(x_k|\breve{Y}^{k-1}, m^{k-1}) \tag{4.7}$$

The prediction density in the above equation is given by the Chapman Kolmogorov equation

$$p(x_k|\breve{Y}^{k-1}, m^{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|\breve{Y}^{k-1}, m^{k-1})dx_{k-1} \tag{4.8}$$

The dynamical equation (4.1) is linear and assuming a Gaussian prior probability density function, we can write $p(x_k|Y^{k-1}, m^{k-1}) = N(x_k; \hat{x}_{k|k-1}, P_{k|k-1})$ where the parameters

$\hat{x}_{k|k-1}$ and $P_{k|k-1}$ are defined as

$$\hat{x}_{k|k-1} = F\hat{x}_{k-1|k-1}$$
$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q$$

Now we need to derive the likelihood part of the Bayesian formula in equation (4.7) for the fusion center. At time k, we receive m_k validated measurements at the sensor node and we assume only one measurement is the target measurement while the rest are clutter measurements. We need data association to calculate probabilities for the origin of each received measurement. There are two possible origins of each measurement, i.e., target and clutter. The clutter measurements are false alarms because sensor (radar) indicate the presence of a target at clutter locations. We define association events $\theta_k(i)$ such that measurement i is a target originated measurement with some probability $p(\theta_k(i))$ where $i = 1, 2..., m_k$. All these association events are mutually exclusive and exhaustive events for each time k. These assumptions are consistent with standard PDA filter assumptions [4]. We utilize the same assumptions for our compressed measurement $\breve{y}_k$ which is the weighted sum of clutter and target measurements. Therefore the likelihood function $p(\breve{y}_k|x_k)$ becomes

$$\begin{aligned} p(\breve{y}_k|x_k) &= \sum_{i=1}^{m_k} p(\breve{y}_k|x_k, \theta_k(i))p(\theta_k(i)) \\ p(\breve{y}_k|x_k, \theta_k(i)) &= \int ... \int p(\breve{y}_k, y_k(j)|x_k, \theta_k(i))dy_j...dy_{m_{k-1}} \end{aligned} \tag{4.9}$$

where j is from the set of m_k validated measurements excluding the i^{th} measurement. The above integral evaluates the likelihood of $\breve{y}_k$ considering the i^{th} measurement in equation (4.4) to be target originated while considering all other measurements to be clutter originated measurements. According to the measurement equation and considering the Gaussian measurement noise assumption, the above equation can be written as

$$p(\breve{y}_k|x_k, \theta_k(i)) = \int ... \int N(\breve{y}_k; \alpha_i Hx_k + \sum_{j=1, j\neq i}^{m-k} \alpha_j y_j, \alpha_i^2 R)dy_j...dy_{m_{k-1}} \tag{4.10}$$

Now the likelihood function can be determined by replacing $p(\breve{y}_k|x_k, \theta_k(i))$ from equation

(4.10) in equation (4.9) and can be written as

$$p(\breve{y}_k|x_k) = \sum_{i=0}^{m_k} \int ... \int N(\breve{y}_k; \alpha_i H x_k + \sum_{j=1, j\neq i}^{m_k} \alpha_j y_j, \alpha_i^2 R) dy_j ... dy_{m_{k-1}} p(\theta_k(i)) \tag{4.11}$$

The formula in equation (4.7) can be written as

$$\begin{aligned}
p(x_k|\breve{Y}_k) &\propto p(\breve{y}_k|x_k) p(x_k|\breve{Y}^{k-1}) \\
p(x_k|\breve{Y}_k) &\propto p(\breve{y}_k|x_k) N(x_k; \hat{x}_{k|k-1}, P_{k|k-1}) \\
p(x_k|\breve{Y}_k) &\propto \sum_{i=1}^{m_k} \int ... \int N(\breve{y}_k; \alpha_i H x_k + \sum_{j=1, j\neq i}^{m_k} \alpha_j y_j, \alpha_i^2 R) \\
&\qquad dy_j ... dy_{m_{k-1}} p(\theta_k(i)) N(x_k; \hat{x}_{k|k-1}, P_{k|k-1}) \\
p(x_k|\breve{Y}_k) &\propto \sum_{i=1}^{m_k} \int ... \int N(\breve{y}_k; \alpha_i H x_k + \sum_{j=1, j\neq i}^{m_k} \alpha_j y_j, \alpha_i^2 R) \\
&\qquad \times N(x_k; \hat{x}_{k|k-1}, P_{k|k-1}) p(\theta_k(i)) dy_j ... dy_{m_{k-1}}
\end{aligned} \tag{4.12}$$

Since, we assume Gaussian prior and likelihood distributions, we need to solve for the product of two Gaussian distributions as can be seen in equation (4.12).

We can solve product of two Gaussian $N(\breve{y}_k; \alpha_i H x_k + \sum_{j=1, j\neq i}^{m_k} \alpha_j y_j, \alpha_i^2 R) N(x_k; \hat{x}_{k|k-1}, P_{k|k-1})$ in above equations using the Gaussian product formula given in [14] which can be written as below using standard Kalman filter notation.

Theorem 4.1. (Gaussian Product Theorem) *For two Gaussian with parameters,* $x_k, \hat{x}_{k|k-1} \in R^{d_1}, H \in R^{d_2 * d_1}, x_k \in R^{d_2}$ *and positive definite matrices* $R, P_{k|k-1}$

$$N(\breve{y}_k : Hx_1, R) N(x_k; \hat{x}_{k|k-1}, P_{k|k-1}) = N(\breve{y}_k; H\hat{x}_{k|k-1}, S_k) N(x_k; \hat{x}_{k|k}, P_{k|k})$$

where

$$
\begin{aligned}
S_k &= HP_{k|k-1}H' + R \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K(\breve{y}_k - H\hat{x}_{k|k-1}) \\
P_{k|k} &= P_{k|k-1} - KHP_{k|k-1} \\
K &= P_{k|k-1}H'S_k^{-1}
\end{aligned}
$$

Using Theorem 4.1, the product of two Gaussian $\beth = N(\breve{y}_k; \alpha_i Hx_k + \sum_{j=1,j\neq i}^{m_k} \alpha_j y_j, \alpha_i^2 R) \times N(x_k; \hat{x}_{k|k-1}, P_{k|k-1})$ in above equations becomes

$$
\begin{aligned}
\beth &= N(\breve{y}_k - \sum_{j=1,j\neq i}^{m_k} \alpha_j y_j; \alpha_i Hx_k, \alpha_i^2 R) \times N(x_k; \hat{x}_{k|k-1}, P_{k|k-1}) \\
&= N(\breve{y}_k - \sum_{j=1,j\neq i}^{m_k} \alpha_j y_j; \hat{\breve{y}}_k, \hat{S}_k) \times N(x_k; x_{k|k}, \hat{P}_{k|K}) \\
&= N(\breve{y}_k; \hat{\breve{y}}_k + \sum_{j=1,j\neq i}^{m_k} \alpha_j y_j, \hat{S}_k) \times N(x_k; \hat{x}_{k|k}, P_{k|k})
\end{aligned}
$$

The parameters of the above Gaussians are calculated using the following equations:

$$
\begin{aligned}
\hat{\breve{y}}_k &= \alpha_i H\hat{x}_{k|k-1} \\
\hat{S}_k &= \alpha_i^2(HP_{k|k-1}H' + R) \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K(\breve{y}_k - \alpha_i H\hat{x}_{k|k} - \sum_{j=1,j\neq i}^{m_k} \alpha_j \breve{y}_j) \\
P_{k|k} &= P_{k|k-1} - \alpha_i KHP_{k|k-1}
\end{aligned}
$$

where $K = \alpha_i P_{k|k-1}H'\hat{S}_k^{-1}$ is the gain of the filter. The posterior probability density function can be written as

$$
\begin{aligned}
p(x_k|Y^k) \propto \sum_{i=1}^{m_k} \int \dots \int N(\breve{y}_k; \hat{\breve{y}}_k + \sum_{j=1,j\neq i}^{m_k} \alpha_j y_j, \hat{S}_k) \times N(x_k; \hat{x}_{k|k}, P_{k|k}) \\
p(\theta_k(i))dy_j \dots dy_{m_{k-1}}
\end{aligned}
\tag{4.13}
$$

4.3.1 Compressed Filter Algorithm

A closed form solution of equation (4.13) is intractable. However, in the literature there exist techniques to solve complex integrals either numerically or using Monte Carlo simulations. We adopt Monte Carlo simulation methods to solve equation (4.13). By adopting a Monte Carlo simulation procedure to find an approximate solution to equation (4.13), we need to simulate $m_k - 1$ variables for each i. When $m_k - 1$ becomes large, the growing computations make this procedure less attractive. We propose another simple approach to handle this computational problem. The summation term $\sum_{j=1, j\neq i}^{m_k} \alpha_j y_j$ is the sum of $m_k - 1$ clutter measurements. Assuming that the distribution of clutter is uniform in the surveillance region, this summation term is actually the sum of uniformly but non-identically distributed random variables. The distribution of the sum of n uniformly distributed random variables over intervals $[a_i, b_i]$ can be found in [10] and expressed in equation (4.14) as

$$\begin{aligned} f_n(s) = \frac{1}{A_n(n-1)!}\{(s^+)^{n-1} + \sum_{v=1}^{n}(-1)^v \sum_{j_1=1}^{n}\sum_{j_2=j_1+1}^{n} \dots \\ \sum_{j_v=j_{v-1}+1}^{n}\{[s-(a_{j_1}+a_{j_2}+\dots+a_{j_v})]^+\}^{n-1}\} \end{aligned} \tag{4.14}$$

However, in our case we have the sum of non-identical but uniformally distributed random variables as written in equation (4.15).

$$S_n = \sum_{j=1}^{n} \alpha_j y_j = \sum_{j=1}^{n} X_j \tag{4.15}$$

The distribution of the new random variable S_n is the convolution of the individual distributions of X_j. Equation (4.16) contains the expression for distribution of resultant random variable S_n however we do not have a closed form solution for this convolution.

$$\begin{aligned} S_n &= X_1 + X_2 + \dots + X_n \\ p_{S_n}(s) &= p_{X_1}(x_1) \star p_{X_2}(x_2) \star \dots \star p_{X_n}(x_n) \\ F_S(s) &= \int_{-\infty}^{s} p_S(u)du \end{aligned} \tag{4.16}$$

Equation (4.13) can be re-written by replacing $\sum_{j=1,j\neq i}^{m_k} \alpha_j y_j$ with another random variable S_n defined above. Now the resultant equation contains one integral and can be easily solved using Monte Carlo integration.

$$p(x_k|\breve{y}_k) \propto \sum_{i=1}^{m_k} \int N(\breve{y}_k; \hat{\breve{y}}_k + S_n, \hat{S}_k) p(S_n) \times N(x_k; \hat{x}_{k|k}, P_{k|k}) p(\theta_k(i)) dS_n \tag{4.17}$$

The parameters of $N(\breve{y}_k; \hat{y}_k + S_n, \hat{S}_k) N(x_k; \hat{x}_{k|k}, P_{k|k})$ are given as.

$$\begin{aligned}
\hat{\breve{y}}_k &= \alpha_i H \hat{x}_{k|k-1} \\
\hat{S}_k &= \alpha_i^2 (H P_{k|k-1} H' + R) \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K(\breve{y}_k - \alpha_i H \hat{x}_{k|k} - S_n) \\
P_{k|k} &= P_{k|k-1} - \alpha_i K H P_{k|k-1}
\end{aligned}$$

where K is the gain of the filter.

The simulation of the algorithm is based on a mixture of Gaussians. We initialize with nG Gaussians and weight and predict each Gaussian using standard prediction as described in equation (4.3). To update each predicted Gaussian and its associated weight, we need to implement equation (4.17). The integral in equation (4.17) is solved by simulating S_n as defined in equation (4.15). Each predicted Gaussian is updated with every sample of S_n and for all data association hypotheses. Then we update the weight of each resultant Gaussian. The number of Gaussians in this mixture grows exponentially with time, however we maintain only the high weighted Gaussians and discard the low weighted Gaussians. The algorithm is summarized in Algorithm 6.

4.4 Simulation Results

Each experiment discussed in this section assumes an identical simulation environment. One target is moving in a straight line in a one dimensional surveillance space. The target is moving with uniform velocity of 2 m/s parallel to the Cartesian x axis. This surveillance space is observed with one sensor with a scan interval of $T = 1s$. Sensor measurement noise is assumed to be Gaussian, white and independent of process noise,

Algorithm 6 : The Compressed Filter one Iteration

Initialize $\leftarrow [w^l_{k-1|k-1}, N(x; \hat{x}^l_{k-1|k-1}, P^l_{k-1|k-1})]^{nG}_{l=1}$
Recursion Starts
for $l = 1$ **to** nG **do**
  Prediction $\leftarrow N(x; \hat{x}^l_{k|k-1}, P^l_{k|k-1})$
  for $i = 1$ **to** m_k **do**
    Draw nS Samples of S_n *as in (4.15)*
    for $j = 1$ **to** nS **do**
      Update $\leftarrow N(x; \hat{x}^{l,i,j}_{k|k}(S_n(j)), P^{l,j,i}_{k|k})$
      $w^{l,i,j}_{k|k} = w^l_{k-1|k-1} N(y_k; \alpha_i H \hat{x}^l_{k|k-1} + S_n(j), \alpha_i^2 S^l_{k|k-1})$
    end for
  end for
end for
Discard Gaussians with low weights
Weighted Sum $\leftarrow \hat{x}_{k|k} = \sum_{i=1}^{G} w^G_{k|k} \hat{x}^G_{k|k}$
Recursion Ends

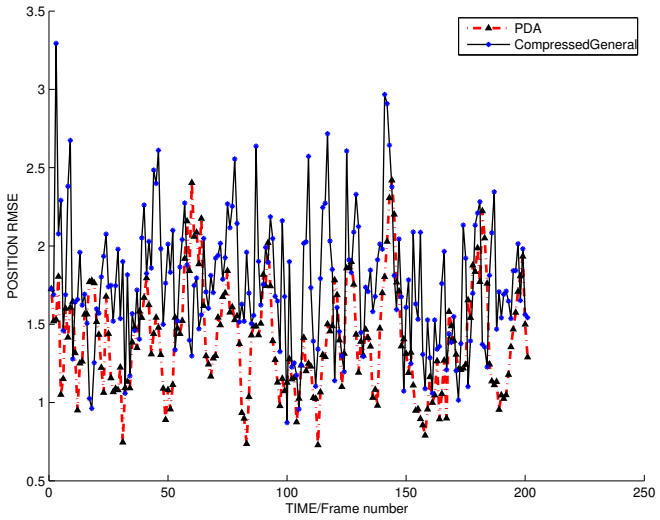

Figure 4.2: The simulations comparison of RMSE of local PDAF filter and proposed global Compressed Filter with very high clutter measurement density of $\rho = 0.1$.

with covariance $R = 4$. The clutter measurement is $0.1/scan/m^2$. The local filter used is a classical probabilistic data association filter [4]. The sensor node is connected to the fusion center via a digital communication channel with an assumption that the network is ideal with no transmission errors and no out of sequence records. The measurements falling inside the PDAF gate are combined into one measurement and the weights are

calculated based on the likelihood of each measurement. The resultant measurements along with the associated scaling factors are sent to the global processing facility. The Compressed Filter utilizes received information to track the target in the surveillance space. Figure 4.2, shows the root mean square error comparison of local PDA filter and the Compressed Filter. The time averaged RMSE values for the local PDA and the CF are 1.6044 and 1.7236 respectively. We see that tracking performance of the CF is slightly worse than PDA but with greatly reduced transmission bandwidth.

4.5 Summary

In this chapter, we proposed a scheme for measurement information transmission in a distributed tracking architecture when clutter is present. This scheme proved to be helpful in transmitting information at low data rates. We combined received measurements at a local node into a single measurement with associated weights, calculated using the likelihood of each measurement. The resultant combined measurement is transmitted to the global processing unit. We proposed a Bayesian recursive filter which operates on the combined measurements. We implemented a Monte carol simulation based method to update the predicted conditional densities. The implementation is based on a Gaussian mixture with associated weights. To limit the number of Gaussians in the mixture, we keep high weighted Gaussians and discard low weighted Gaussians. The simulation results show that tracking performance is slightly reduced but good tracking is maintained. Our approach requires lower communication bandwidth and maintains the track and its quality. It is worth to note that the encoding of data with prior distribution requires lower number of bits as compared to data encoding with no prior distribution. This is preliminary work and further work is needed to study practical tracking limitations such as missing measurements, probability of detection is less than 1, incorporating encoding and higher dimension cases.

Chapter 5
Particle Based Distributed Estimation

This chapter describes a novel idea for distributed estimation and tracking based on transmitting samples of the local posterior distribution to the fusion center. The local sensor node performs tracking/estimation and updates the posterior probability distribution function of the state with every new received sensor measurement. The local node is connected to the fusion center via a low data rate communication channel. In the proposed method, we transmit only one random sample or particle of the local posterior probability density function to the fusion center. At the fusion center, the particles are used to develop a global posterior probability distribution function (pdf). We calculate the weight of the most recent particle and update the weights of previously received particles using adaptive importance sampling techniques. The most recent samples will have higher weights due to the closely matched target and sampling distributions when compared with older samples. As we progress in time, the number of the received particles grows and the estimates improve. We use the concept of adaptive importance sampling because the received samples upto time k have different importance sampling distributions and their target distribution keeps changing with time. We study the statistical properties of the weights of the samples such as mean and variances of the weights and their contribution to the convergence of the estimate. Most of the computations at the fusion center are required to update the weight of the samples received, however following the iterative weight update procedure, these can be reduced. This chapter focuses on the simplest case of parameter estimation, however this idea can be easily generalised to the dynamic case of target tracking and multiple sensors.

5.1 Introduction

IN Chapter 3, we discussed transmitting scaled equivalent measurements with low data rates using distributed tracking methods without considering clutter in the surveillance region. The Compressed Filter based on using a weighted sum of clutter and target measurements as a virtual measurement is proposed in Chapter 4. In both

contributions, we transmit forms of measurements or processed measurements to the fusion center. In practice many distributed tracking systems transmit the moments of the local posterior distribution. For example in the case of Kalman filtering, transmitting only the first two moments of the posterior distribution to the fusion center represents the complete distribution because of Gaussianity. In the case of non-linear estimation transmitting the first two moments is not enough to reproduce the optimal global posterior distribution. The proposed approach is based on sampling theory of distributions, i.e., if we have enough samples of one distribution then we can sub-optimally reproduce the the original distribution [58]. Importance Sampling (IS) is the technique used to approximate the target distribution when samples are drawn from another distribution [58].

Monte Carlo approximation methods have received significant attention in target tracking and estimation due to their ability to handle non-linear systems and because they are computationally practical. Particle filters are the most famous among sequential Monte Carlo methods and have been in the spot-light for tracking following the pioneering work of Gordon et al. in 1993 [27]. We have discussed the background to Monte Carlo simulation based tracking algorithms in Chapter 2. Thinking along the same lines, in this chapter we propose a distributed estimator employing Monte Carlo simulation.

We consider distributed estimation/tracking based on particles transmission and propose a global estimator which requires only one sample drawn from the local posterior distribution at each time. We consider that tracking is being performed both locally (sensor node) and remotely (fusion center). The local tracker uses sensor measurements and maintains the local track state. We draw random samples from the local track state posterior distribution and transmit them to the fusion center (remote location). At the fusion center, we treat these samples as virtual measurement information and update the track state using these samples. We utilize the concept of importance sampling and calculate the weight of each received sample. The weights of previously received samples are updated using new information. In the proposed method, we transmit only one sample to the fusion center. The process is explained below in detail.

At time k the central node has received k samples randomly drawn from the local posterior distributions of time 1 to time k. The target distribution is $\pi_k(x) = p(x_k|Y^k)$

and the sampling distributions of the samples $X_1, X_2, X_3, ..., X_k$ are $\pi_1(x)$, $\pi_2(x)$, $\pi_3(x)$, ..., $\pi_k(x)$ respectively. Now using importance sampling the weights of each sample can be calculated by the ratio of its target and sampling distributions. So, we can easily see that every sample has a different sampling distribution while having the same target distribution. Therefore, this problem can be viewed as adaptive importance sampling [66]. As time progresses, the target density changes, the number of samples increases by one and the weight of each sample needs to be modified. We can update the weights of the received samples iteratively using adaptive importance sampling techniques since the target distribution is changing with time k and every sample has its own sampling distribution. Given k samples and their respective weights we can approximate the expected value of the posterior distribution at the fusion center by finding the weighted sum of all received samples. In this chapter, we restrict ourselves to the static parameter estimation problem in order to develop convergence results but this work can be extended in a straight forward manner to the more general dynamic case. In order to study the efficiency of any estimator, estimation error and consistency are very important factors to study. Convergence is studied in the next chapter.

The rest of this chapter discusses the proposed Particle Based Distributed Filter (PBDF). Section 2 formulates the problem of parameter estimation. We discuss the proposed filter in section 3 and its properties in section 4. We also discuss the properties of the weights of the samples and the mean and variance of the proposed estimator in section 4. Simulation results and aspects of convergence are studied in section 5. A rigorous convergence result will be covered in the next chapter.

5.2 Problem Formulation

As a simple case, we wish to estimate a time invariant parameter x using noisy measurements. The measurement model is defined by

$$y_k = x + \omega_k \tag{5.1}$$

where ω_k is measurement noise which is zero mean white Gaussian noise with variance R. The Bayesian estimate of the parameter x can be written as

$$p(x|Y^k) = \frac{p(y_k|x)p(x|Y^{k-1})}{\int p(y_k|x)p(x|Y^{k-1})dx} \tag{5.2}$$

where Y^k is the set of all measurements received upto time k.

We consider a single sensor node which is connected to a central processing unit via a forward only communication link. We wish to find the global posterior pdf by receiving information extracted from the local posterior pdf.

We aim to solve the distributed estimation problem by randomly drawing one sample at each time step from the local posterior distribution, and sending it to the fusion center. At the fusion center, we calculate or update weights for all received particles upto the current time and our global estimator is defined as a weighted sum of the received samples. All the derivations are based on the static parameter estimation case and the dynamic case is left for future work.

5.3 Proposed Distributed Estimator

We draw sample X_k from the local posterior pdf $p^L(x|Y^k)$ at each time k and transmit it to the fusion center. At the fusion center, we receive this sample and calculate its weight w_k and update the weights for all previously received samples. We reconstruct the posterior pdf conditioned on received samples and measurements at the fusion center $p(x|X^k, Y^k)$ with the help of received samples and their associated updated weights.

The calculation of the weights of the samples uses importance sampling concepts found in [59]. As time increases, the number of received samples increases leading to increasingly better estimates of the parameter. The posterior density based on previously received samples changes with time therefore we need to update the weight of every received sample at the current time. Let $\pi_k(x) = p^L(x|Y^k)$ denotes the local posterior pdf at time k. We derive the proposed estimator and include the procedure to update the sample weights as below.

Our estimator is a weighted sum of the received samples but the major question is

how these weights are calculated and what is their role in the convergence of the estimator. The mechanism used for calculation of the weights of a sample is adaptive importance sampling which means that all received samples have different importance sampling distributions. Also, as time progresses the target density will change. Due to this two way adaptive importance sampling, the study of the convergence of the proposed filter is challenging.

We derive the PBDF in the conventional way using adaptive importance sampling and an induction procedure. The samples are denoted by X_τ where τ represents the time of drawing and/or transmitting while the weights are denoted as w_τ^o where τ denotes the time at which sample is drawn and o denotes the time of the update.

Starting with time 1, we have just one sample and $w_1 = 1$ since the target and importance densities are the same,

$$\begin{aligned} X_1 &\sim \pi_1(x) \\ w_1^1 &\propto 1 \end{aligned} \tag{5.3}$$

For $k = 2$, there are two samples X_1 and X_2 whose weights are updated as below,

$$\begin{aligned} X_2 &\sim \pi_2(x) \\ w_2^2 &\propto 1 \end{aligned} \tag{5.4}$$

The weight of the previously received sample X_1 is updated considering $\pi_2(x)$ as the target density and $\pi_1(x)$ as the sampling density.

$$\begin{aligned} w_1^2 &= \frac{\pi_2(X_1)}{\pi_1(X_1)} \\ &= \frac{l_2(X_1) l_1(X_1) \pi_0(X_1)}{\int l_2(X_1) l_1(X_1) \pi_0(X_1) dX_1} \frac{\int l_1(X_1) \pi_0(X_1) dX_1}{l_1(X_1) \pi_0(X_1)} \\ &= \frac{l_2(X_1)}{\int l_2(X_1) l_1(X_1) \pi_0(X_1) dX_1} \frac{\int l_1(X_1) \pi_0(X_1) dX_1}{1} \\ &= l_2(X_1) \frac{\int l_1(X_1) \pi_0(X_1) dX_1}{\int l_2(X_1) l_1(X_1) \pi_0(X_1) dX_1} \\ &= l_2(X_1) \frac{C_1^1}{C_1^2} \end{aligned} \tag{5.5}$$

where $l_2(X_1) = p(y_2|X_1)$, is the likelihood of measurement y_2 with respect to the sample X_1, $C_1^1 = \int l_1(X_1)\pi_0(X_1)dX_1$ and $C_1^2 = \int l_2(X_1)l_1(X_1)\pi_0(X_1)dX_1$. The proposed estimator using two available samples can be written as

$$\Delta_2 = \frac{w_1^2 X_1 + w_2^2 X_2}{2} \tag{5.6}$$

For $k = 3$, there are three samples $X1$, X_2 and X_3 drawn from $\pi_1(x)$, $\pi_2(x)$ and $\pi_3(x)$ respectively and all have the same target density $\pi_3(x)$. The associated weights of the currently available samples are calculated using

$$\begin{aligned} X_3 &\sim \pi_3(x) \\ w_3^3 &\propto 1 \end{aligned} \tag{5.7}$$

The weight of the sample received at $k = 1$ is updated below. For sample X_1, the importance density is $\pi_1(x)$ and target density is $\pi_3(x)$.

$$\begin{aligned} w_1^3 &= \frac{\pi_3(X_1)}{\pi_1(X_1)} \\ &= \frac{l_3(X_1)l_2(X_1)l_1(X_1)\pi_0(X_1)}{\int l_3(X_1)l_2(X_1)l_1(X_1)\pi_0(X_1)dX_1} \frac{\int l_1(X_1)\pi_0(X_1)dX_1}{l_1(X_1)\pi_0(X_1)} \\ &= \frac{l_3(X_1)l_2(X_1)}{\int l_3(X_1)l_2(X_1)l_1(X_1)\pi_0(X_1)dX_1} \frac{\int l_1(X_1)\pi_0(X_1)dX_1}{1} \\ &= l_3(X_1)l_2(X_1) \frac{\int l_1(X_1)\pi_0(X_1)dX_1}{\int l_3(X_1)l_2(X_1)l_1(X_1)\pi_0(X_1)dX_1} \\ &= l_3(X_1)l_2(X_1) \frac{C_1^1}{C_1^3} \end{aligned} \tag{5.8}$$

Using the equation (5.5), we can write the expression for w_1^3 as a function of w_1^2 as

$$w_1^3 = l_3(X_1) w_1^2 \frac{C_1^2}{C_1^3} \tag{5.9}$$

where $l_3(X_1) = p(y_3|X_1)$, $C_1^2 = \int l_2(X_1)l_1(X_1)\pi_0(X_1)dX_1$ and $C_1^3 = \int l_3(X_1)l_2(X_1)l_1(X_1)\pi_0(X_1)dX_1$. The weight of the sample received at time $k = 2$ is updated using equation (5.10). For the sample X_2, the importance density is $\pi_2(x)$ and

the target density is $\pi_3(x)$

$$\begin{aligned} w_2^3 &= \frac{\pi_3(X_2)}{\pi_2(X_2)} \\ &= \frac{l_3(X_2)l_2(X_2)l_1(X_2)\pi_0(X_2)}{\int l_3(X_2)l_2(X_2)l_1(X_2)\pi_0(X_2)dX_2} \\ &\times \frac{\int l_2(X_2)l_1(X_2)\pi_0(X_2)dX_2}{l_2(X_2)l_1(X_2)\pi_0(X_2)} \\ &= l_3(X_2)\frac{\int l_2(X_2)l_1(X_2)\pi_0(X_2)dX_2}{\int l_3(X_2)l_2(X_2)l_1(X_2)\pi_0(X_2)dX_2} \\ &= l_3(X_2)\frac{C_2^2}{C_2^3} \\ &= l_3(X_2)w_2^2\frac{C_2^2}{C_2^3} \end{aligned} \tag{5.10}$$

where $l_3(X_2) = p(y_3|X_2)$, $C_2^2 = \int l_2(X_2)l_1(X_2)\pi_0(X_2)dX_2$ and $C_2^3 = \int l_3(X_2)l_2(X_2)l_1(X_2)\pi_0(X_2)dX_2$. Using three samples the estimate becomes

$$\Delta_3 = \frac{w_1^3X_1 + w_2^3X_2 + w_3^3X_3}{3} \tag{5.11}$$

Following the above procedure for weight updates, we can write expressions for weight update of $X_1^k, X_2^k, ..., X_k^k$ as .

$$\begin{aligned} w_1^k &= l_k(X_1)w_1^{k-1}\frac{C_1^{k-1}}{C_1^k} \\ w_2^k &= l_k(X_2)w_2^{k-1}\frac{C_2^{k-1}}{C_2^k} \\ w_3^k &= l_k(X_3)w_3^{k-1}\frac{C_3^{k-1}}{C_3^k} \\ &. \\ &. \\ &. \\ w_{k-1}^k &= l_k(X_{k-1})w_{k-1}^{k-1}\frac{C_{k-1}^{k-1}}{C_{k-1}^k} \\ w_k^k &\propto 1 \end{aligned} \tag{5.12}$$

where $l_k(X_n) = p(y_k|X_n)$, $C_n^k = \int l_k(X_n)l_{k-1}(X_n)...l_1(X_n)\pi_0(X_n)dX_n$ and $n = 1, 2, ..., k$.

The above results are derived using target density $\pi_k(x)$ and respective importance densities of each sample. Generally, a sample is drawn at time n with importance density $\pi_n(x)$ and target density $\pi_k(x)$ then its weight w_n^k is calculated using the ratio $\pi_k(x)/\pi_n(x)$. The general form of the estimator at time k using k available samples is

$$\Delta_k = \frac{w_1^k X_1 + w_2^k X_2 + w_3^k X_3 + ... + w_k^k X_k}{k} \tag{5.13}$$

Equation (5.13) is the expected value of the parameter of interest calculated at the fusion center while the weights are defined in equation (5.12). From equation (5.12), we can observe that the most recent samples have higher weights while the old samples carry low weights. This is supported by the fact that the most recent samples have lower distances in their target and importance densities. For a very large k, the posterior pdf of the parameter does not change much and we call this the steady state distribution. Based on this intuitive argument, if we have sample drawn from $\pi_{k-1}(x)$ then the weight w_{k-1}^k of sample X_{k-1} approaches 1.

The PBDF algorithm is straightforward to implement and does not depend on the type of the local estimator being implemented at the sensor node. The only information we need from the local end is a sample of the posterior pdf. The local end posterior pdf can be of any shape and the local tracker can be linear or non-linear. The sample is drawn randomly from the local posterior pdf and transmitted to the fusion center. The PBDF is outlined in Algorithm 7. For the sake of simplicity, we use a classical Kalman filter at the local end.

In order to study the performance of PBDF, we need to look at the estimation error and consistency. The Root Mean Square Error (RMSE) is compared with the optimal local Kalman filter in the simulation section of this chapter. In order to study consistency, we examine the asymptotic behavior of the expected value of Δ_k and the variance of Δ_k. As a necessary condition for convergence, we need to show that the mean and variance of the proposed estimator remain bounded for large k. From the weighted sum equation, we see that the statistics of the proposed PBDF depend on the calculation of the weights. In other words, if our weights stay bounded and do not keep increasing with time then

the proposed estimator stays bounded. Therefore in order to study the statistics of the proposed estimator, we study the dynamic behavior of the weights associated with the samples. In the next section, we study the properties of the proposed PBDF including statistics of weight samples.

5.4 Properties of the Proposed Estimator

We defined the proposed estimator at the fusion center which operates on using received samples one at a time from the sensor node. The weight of each received sample is updated and the proposed estimator is defined by a weighted sum of all received samples until time k. The formula of this estimator is written in equation (5.13) and clearly shows its dependence on the weights of the samples. In this section, we study the properties of both the proposed estimator and the weights of the received samples. We derive expressions for estimator variance as a function of mean and variance of the weights. Since, mean and variance of the proposed estimator depend on the mean and variance of the weights, we study the weights statistics first.

5.4.1 Weights Statistics

In this subsection, we derive and discuss the mean and variance of the weights of the samples at time k. The weights of each sample can be considered as a random quantity and we want to know the asymptotic behavior of the weights. Intuitively, we expect the most recent samples will have larger weights (on average) as compared to the older samples. This argument makes sense in that as time passes we have more and more measurements at the local end which leads to a better estimation. Therefore, the posterior pdf tends towards a steady state value and samples taken from most recent local posterior pdf contain more information. Since, we are dealing with randomly drawn samples the results should be averaged when compared. Let $\hat{\phi}_n^k = E[w_n^k|Y^k]$ denotes the conditional mean of the weight of sample drawn at time n and updated at time k. Similarly, the variance of the weight of the sample drawn at time n and updated at time k is denoted as Ψ_n^k. The mean of w_1^k is calculated using

Algorithm 7 : The Particle Based Distributed Filter

Local End
for $n = 1$ **to** k **do**
Prediction Step

$$\hat{x}_{n|n-1} = \hat{x}_{n-1|n-1}$$
$$P_{n|n-1} = P_{n-1|n-1}$$

Compute Kalman gain K_n and predicted measurement $\hat{y}_{n|n-1}$ and covariance of innovations

$$K_n = P_{n|n-1}H^T S_n^{-1}$$
$$\hat{y}_{n|n-1} = H\hat{x}_{n|n-1}$$
$$S_n = HP_{n|n-1}H^T + R$$

Update step with measurement y_k

$$\hat{x}_{n|n} = \hat{x}_{n|n-1} + K_k(y_n - H\hat{x}_{n|n-1})$$
$$P_{n|n} = (I - K_nH)P_{n|n-1}(I - K_nH)^T + K_nRK_n^T$$

Resulted posterior pdf $p(x_n|Y^n) = N(x_n; H\hat{x}_{n|n}, P_{n|n})$
Draw random sample X_n from posterior pdf $N(x_n; H\hat{x}_{n|n}, P_{n|n})$
Transmit X_n to the fusion center
end for
At the Fusion Center
for $n = 1$ **to** k **do**
Receive sample X_n every time
Calculate the weight of the most recent sample X_n using equation (5.12)
Update the weights of the previously received samples using equation (5.12)
The expected value of the PBDF at the fusion center is calculated using below equation

$$\Delta_n = \sum_{i=1}^{n} \frac{w_i^n X_i}{n}$$

end for
Return the expected value of global estimator as Δ_n

$$\begin{aligned}\hat{\phi}_1^k &= E\left[\frac{\pi_k(X_1)}{\pi_1(X_1)}\right] \\ &= \int \frac{\pi_k(X_1)}{\pi_1(X_1)} p(X_1) dX_1\end{aligned} \tag{5.14}$$

Since the distribution of X_1 is $\pi_1(x)$ the above expression can be written as

$$\begin{aligned}\hat{\phi}_1^k &= \int \frac{\pi_k(X_1)}{\pi_1(X_1)} p(X_1) dX_1 \\ &= \int \frac{\pi_k(X_1)}{\pi_1(X_1)} \pi_1(X_1) dX_1 \\ &= \int \pi_k(X_1) dX_1 \\ &= 1\end{aligned} \tag{5.15}$$

Similarly, we can derive a general expression for $\hat{\phi}_n^k$, where $n = 1, 2, 3, ..., k$ and is given in below set of equations.

$$\begin{aligned}\hat{\phi}_n^k &= E\left[\frac{\pi_k(X_n)}{\pi_n(X_n)}\right] \\ &= \int \frac{\pi_k(X_n)}{\pi_n(X_n)} p(X_n) dX_n\end{aligned} \tag{5.16}$$

Since the distribution of X_n is $\pi_n(x)$ the above expression can be written as

$$\begin{aligned}\hat{\phi}_n^k &= \int \frac{\pi_k(X_n)}{\pi_n(X_n)} p(X_n) dX_n \\ &= \int \frac{\pi_k(X_n)}{\pi_n(X_n)} \pi_n(X_n) dX_n \\ &= \int \pi_k(X_n) dX_n \\ &= 1\end{aligned} \tag{5.17}$$

The expected values of conditional weights $w_1^k, w_2^k, w_3^k ...., w_{k-1}^k$ are all equal to 1 and it is straightforward to see that the mean of the weights remains bounded and is always equal to one. The respective variances of the weights can be calculated using generalized expressions below. The detailed derivation of the variance of the weights is given in

Appendix A.2.

$$\Psi_1^k = \frac{1}{(4\pi R)^{\frac{k-1}{2}}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \prod_{i=3}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_{i-1})} - 1$$

$$\Psi_2^k = \frac{1}{(4\pi R)^{\frac{k-2}{2}}} \frac{N(y_3; H\hat{x}_{2|2}, S_2 - R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \prod_{i=4}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_{i-1})} - 1$$

$$\Psi_3^k = \frac{1}{(4\pi R)^{\frac{k-3}{2}}} \frac{N(y_4; H\hat{x}_{3|3}, S_3 - R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \prod_{i=5}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_{i-1})} - 1$$

.

.

.

$$\Psi_{k-1}^k = \frac{1}{(4\pi R)^{\frac{1}{2}}} \frac{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1} - R/2)}{N^2(y_k; H\hat{x}_{k-1|k-1}, S_{k-1})} - 1$$

Using the above expressions, the general form of the variance of the weights can be calculated using induction

$$\begin{aligned} \Psi_n^k = & \frac{1}{\sqrt[k-n]{4\pi R}} \frac{N(y_{n+1}; H\hat{x}_{n|n}, S_n - R/2)}{N^2(y_{n+1}; H\hat{x}_{n|n}, S_n)} \\ & \times \prod_{i=n+2}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} - 1 \end{aligned} \tag{5.18}$$

Where $S_{k-1} = HP_{k|k-1}H' + R$, $\bar{S}_{i-1} = H\bar{P}_{i|i-1}H' + R/2$ and $n \leqslant k$ represents the time at which samples were drawn and k is current time index.

In summary, the expected value of the weight of the sample drawn at time n and updated at time k is 1 and variance is written in equation (5.18).

The general expression for Ψ_n^k is complex and it is difficult to explain the meaning of this expression. In order to make intuitive arguments about the variance of the weights, we write the recursive form of Ψ_n^k. The set of equations below denotes Ψ_n^k as a function of Ψ_n^{k-1}. We include the derivation for the variance of the weights $Var(w_1^k|Y^k)$ in recursive

form and then write the general form using the same procedure.

$$\begin{aligned}\Psi_1^2 &= \frac{1}{\sqrt{4\pi R}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} - 1 \\ &= \sqrt{\frac{4\pi S_1}{4\pi R}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N(y_2; H\hat{x}_{1|1}, S_1/2)} - 1 \\ &= \sqrt{\frac{S_1}{R}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N(y_2; H\hat{x}_{1|1}, S_1/2)} - 1\end{aligned} \tag{5.19}$$

The recursive expression for Ψ_1^3 can be written as

$$\begin{aligned}\Psi_1^3 &= \sqrt{\frac{1}{4\pi R}}\sqrt{\frac{S_1}{R}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N(y_2; H\hat{x}_{1|1}, S_1/2)} \\ &\times \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y3; H\hat{x}_{2|2}, S2)} - 1 \\ &= \sqrt{\frac{4\pi S_2}{4\pi R}}\sqrt{\frac{S_1}{R}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N(y_2; H\hat{x}_{1|1}, S_1/2)} \\ &\times \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N(y3; H\hat{x}_{2|2}, S_2/2)} - 1 \\ &= \sqrt{\frac{S_2}{R}}\frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N(y3; H\hat{x}_{2|2}, S_2/2)} \times \left[\Psi_1^2 + 1\right] - 1\end{aligned} \tag{5.20}$$

Using the above procedure, we can easily write the recursive form of Ψ_n^k as.

$$\begin{aligned}\Psi_1^k &= \sqrt{\frac{S_{k-1}}{R}}\frac{N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{k-1|k-1}H' + R/2)}{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1}/2)} \\ &\times\left\{\Psi_1^{k-1} + 1\right\} - 1 \\ &= \sqrt{T_1(k)}T_2(k)\{\Psi_1^{k-1} + 1\} - 1 \\ &= T_k\{\Psi_1^{k-1} + 1\} - 1\end{aligned} \tag{5.21}$$

where $T_1(k) = S_{k-1}/R$ and $T_2(k) = N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{k-1|k-1}H' + R/2)/N(y3; H\hat{x}_{k-1|k-1}, S_{k-1}/2)$. Similarly, we derive the recursive form of the variance

of w_n^k.

$$\begin{aligned}\Psi_n^k &= \sqrt{\frac{S_{k-1}}{R}}\frac{N(y_k; H\bar{x}_{k-1|k-1}, \bar{S}_{k-1})}{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1}/2)} \times \left\{\Psi_n^{k-1} + 1\right\} - 1 \\ &= \sqrt{T_1(k)T_2(k)}\{\Psi_n^{k-1} + 1\} - 1 \\ &= T_k\{\Psi_n^{k-1} + 1\} - 1\end{aligned} \tag{5.22}$$

The parameters involved in above set of equations are

$$\begin{aligned}\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - K_k H P_{k|k-1} \\ \bar{x}_{k|k} &= \bar{x}_{k|k-1} + \bar{K}_k(y_k - H\bar{x}_{k|k-1}) \\ \bar{P}_{k|k} &= \bar{P}_{k|k-1} - \bar{K}_k H \bar{P}_{k|k-1}\end{aligned} \tag{5.23}$$

where $K_k = P_{k|k-1}H'S_{k-1}^{-1}$ and $\bar{K}_k = \bar{P}_{k|k-1}H'\bar{S}_{k-1}^{-1}$ are the gains for different measurement noise variances.

The recursive form of the variance of the weights is given in equation (5.22). An intuitive argument about the convergence of the sequence Ψ_n^k can be easily drawn using equation (5.22) for all instances of n. The term T_k plays a very important role in the convergence of the sequence Ψ_n^k and it has to converge to 1. From the expression for T_k, the parameters of the Gaussians in the numerator and denominator converge to the same value by utilizing knowledge of Kalman filter convergence and hence this ratio goes to one asymptotically. The value of S_{k-1} converges to R and hence $\sqrt{S_{k-1/R}}$ converges to 1. Based on this brief explanation, we can expect the value of T_k to converge to 1 and therefore the sequence Ψ_n^k converges for all values of n as shown by simulations. A rigorous and complete proof of convergence is presented in the next chapter.

5.4.2 Statistics of the Proposed Estimator

In this subsection, we study the mean and variance statistics of the proposed estimator as defined in equation (5.13). As a necessary condition for the estimator to be consistent, the mean and variance of the proposed estimator must be bounded and variance tends

to zero. We assume that samples are drawn independently. As a first step, we derive expressions for the conditional mean of the proposed PBDF and then we include variance expressions. We let $\hat{\theta}_n^k = E[w_n^k X_n | Y^k]$ and Θ_n^k the conditional mean and variance of the estimator. The subscript n denotes the sampling time while superscript k denotes the updating time of the samples in the weight expressions. The expected value of the proposed estimator defined in equation (5.13) is

$$\begin{aligned} E[\Delta_k] &= E\left[\frac{w_1^k X_1 + w_2^k X_2 + w_3^k X_3 + ... + w_k^k X_k}{k}\right] \\ &= \frac{1}{k}\Big\{E\left[w_1^k X_1 + w_2^k X_2 + w_3^k X_3 + ... + w_k^k X_k\right]\Big\} \\ &= \frac{1}{k}\Big\{E[w_1^k X_1] + E[w_2^k X_2] + E[w_3^k X_3] + ... + E[w_k^k X_k]\Big\} \\ &= \frac{1}{k}\Big\{\int X_1 \pi_k(X_1) dX_1 + \int X_2 \pi_k(X_2) dX_2 \\ &+ \int X_3 \pi_k(X_3) dX_3 + ... + \int X_k \pi_k(X_k) dX_k\Big\} \\ &= \hat{x}_{k|k} \end{aligned} \tag{5.24}$$

According to the above equation the expected value of the proposed estimator is exactly the same as the expected value of the local optimal estimator however the variance needs to be studied before making any claims about the consistency of the proposed method. The variance of the estimator defined in equation (5.13) is

$$Var[\Delta_k] = Var\left[\frac{w_1^k X_1 + w_2^k X_2 + w_3^k X_3 + ... + w_k^k X_k}{k}\right] \tag{5.25}$$

By using the theorem on the variance of the sum of independent random variables given in [70], we can write

$$\begin{aligned} Var[\Delta_k] &= \frac{1}{k^2}\left[\Theta_1^k + \Theta_2^k + \Theta_3^k + ... + \Theta_{k-1}^k + \Theta_k^k\right] \\ &= \frac{1}{k^2}\sum_{n=1}^{k}\Theta_n^k \end{aligned} \tag{5.26}$$

Analytical expressions for the variances of $w_n^k X_n$ from $n = 1$ to $n = k$ are given below. A detailed derivation of these expressions is included in Appendix A.5. We include

results of $Var(w_1^k X_1|Y^k)$ in below set of equations and similarly we can write general expressions for $Var(w_n^k X_n|Y^k)$ using an induction afterwards.

$$\begin{aligned}
\Theta_1^2 &= \frac{1}{\sqrt{4\pi R}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\
&\quad \times \left\{ \bar{P}_{2|2} + \bar{x}_{2|2}^2 \right\} - \hat{x}_{2|2}^2 \\
\Theta_1^3 &= \frac{1}{4\pi R} \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \\
&\quad \times \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \left\{ \bar{P}_{3|3} + \bar{x}_{3|3}^2 \right\} - \hat{x}_{3|3}^2 \\
\Theta_1^4 &= \frac{1}{\sqrt[3]{4\pi R}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\
&\quad \times \prod_{i=3}^{4} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \left\{ \bar{P}_{4|4} + \bar{x}_{4|4}^2 \right\} - \hat{x}_{4|4}^2
\end{aligned} \tag{5.27}$$

Similarly, following the above expression, we can write an expression for Θ_1^{k-1} as below.

$$\begin{aligned}
\Theta_1^{k-1} &= \frac{1}{\sqrt[k-2]{4\pi R}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\
&\quad \times \prod_{i=3}^{k-1} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \left\{ \bar{P}_{k-1|k-1} + \bar{x}_{k-1|k-1}^2 \right\} - \hat{x}_{k-1|k-1}^2 \\
\Theta_1^{k} &= \frac{1}{\sqrt[k-1]{4\pi R}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\
&\quad \times \prod_{i=3}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2
\end{aligned} \tag{5.28}$$

The general expression of $Var(w_n^k X_n|Y^k)$ is

$$\begin{aligned}
\Theta_n^k &= \frac{1}{\sqrt[k-2]{4\pi R}} \frac{N(y_{n+1}; H\hat{x}_{n|n}, S_n - R/2)}{N^2(y_{n+1}; H\hat{x}_{n|n}, S_n)} \prod_{i=n+2}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\
&\quad \times \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2
\end{aligned} \tag{5.29}$$

As a computational load, the above equation involves $k - n - 1$ products of Gaussian divisions. We can write Θ_n^k in recursive form as a function of Θ_n^{k-1}. The derivation of the

expression in equation (5.30) is given in Appendix A.6.

$$\Theta_n^k = \sqrt{\frac{S_{k-1}}{R}} \frac{N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{k-1|k-1}H' + R/2))}{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1}/2)} \left\{ \frac{\Theta_n^{k-1} + \hat{x}_{k-1|k-1}^2}{\bar{P}_{k-1|k-1} + \bar{x}_{k-1|k-1}^2} \right\} \times \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2 \tag{5.30}$$

The general expression for the variance of the proposed estimator is

$$\begin{aligned} Var[\Delta_k] &= \frac{1}{k^2} \sum_{n=1}^{k} \Theta_n^k \\ &= \frac{1}{k^2} \sum_{n=1}^{k} \frac{1}{(4\pi R)^{k-n/2}} \frac{N(y_{n+1}; H\hat{x}_{n|n}, S_n - R/2)}{N^2(y_{n+1}; H\hat{x}_{n|n}, S_n)} \prod_{i=n+2}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\ &\times \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2 \end{aligned} \tag{5.31}$$

The expression in equation (5.31) represents the variance of the proposed estimator and it is difficult to comment on its convergence. However, one obvious aspect is that if the individual variances Θ_n^k for all $n = 1, 2, ..., k$ converge then their sum also converges. Also, this variance depends on the variance of the weights and its convergence depends on the convergence of the weight variances. A detailed analysis of the variance of the proposed estimator is included in the next chapter.

5.4.3 Simulations of the Statistics of the Weights

In this subsection, we simulate the variance of the weights and discuss these results. In our expressions of the the variance of the weights, we made an intuitive argument that if T_k converges to 1 then the variance of our weights converges. Therefore we include the simulations of T_k defined in equation (5.22). The term T_k can be written in terms of two sub-terms $T_1(k)$ and $T_2(k)$. We simulate the first of these sub-terms $T_1(k)$ and $T_2(k)$ and then their complete product T.

Figure 5.4 shows simulations of the term T_k for $k = 15000$ and it is evident from the figure that the value converges to 1. For higher value of n, this convergence is more rapid and is consistent with our theoretical results. Based on this result, we can say that the

term T_k involved in equation (5.22) converges to 1 for all values of n and large k.

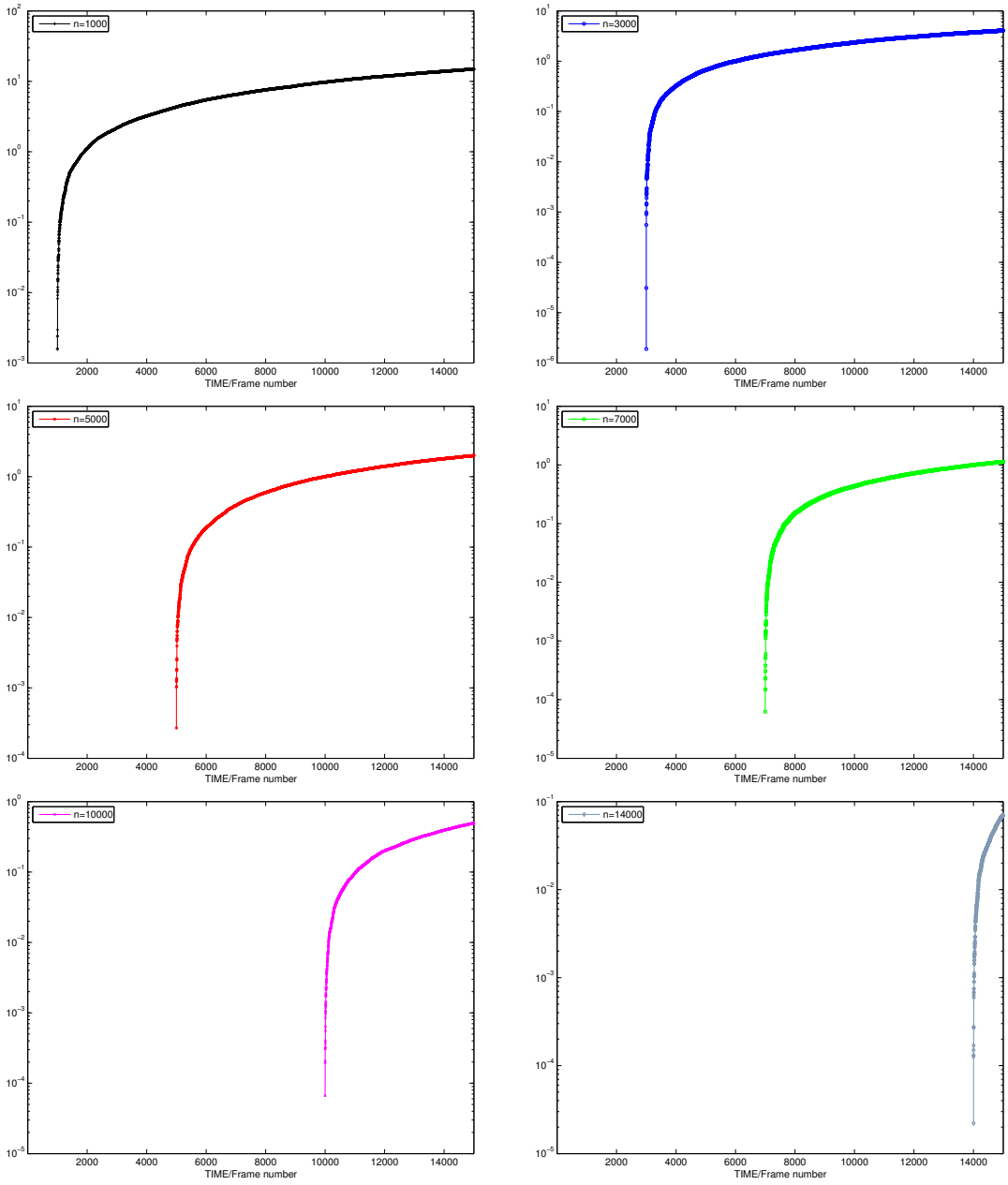

Figure 5.1: Simulation results of the convergence of the individual variance of the the weights with $n = 1000, 3000, 5000, 7000, 10000, 14000$ and $k = 15000$. The vertical axis shows the logarithmic values of the individual variances.

Now, we include simulation of the individual variance of the weights Ψ_n^k for $n = 1000, 3000, 5000, 7000, 10000, 14000$ and $k = 15000$. Figure 5.1 represents the variance of

the weights for different values of n and it is shown from these sub-figures that the variance converges to some steady state value however the converges rate is slow. The x-axes are in linear scale while log scale is used for y-axes to make representation obvious. These are the individual sequences of the variance of the weights of the samples drawn at previous times and updated gradually with the increase of the time. We also observe very small change of the weight variance as k gets larger. The convergence of the weight variance depends only on large value of k and is independent of the value of n. We can have similar convergence results for all n starting from 1 to the value of k.

Simulation results presented in this section suggest the convergence of the sequence of the weights variance and hence contribute to the convergence of the proposed PBDF with a low convergence rate. In the next section, we discuss the individual variances of Θ_n^k.

5.4.4 Simulations of the Statistics of the PBDF

In this subsection, simulations of the individual variances of the proposed estimator are included. We simulate each Θ_n^k for $n = 1000, 3000, 5000, 7000, 10000$ and $k = 15000$ and include our results. Again, the x-axis is in linear scale while the y-axes is represented in log-scale.

From the Figure 5.2, it is shown that the term Θ_n^k for n converges to some value and in-fact this is true for all values of n. Therefore, all the the terms in equation (5.31) inside the summation converge and therefore the $Var[\Delta_k]$ converges.

5.5 Simulations of the Proposed PBDF

In this subsection, we summarize the simulation results of variances of the weights and estimator along with simulation of the proposed estimator. A standard Kalman filter processes the sensor data and every time we transmit only one random sample drawn from the local posterior pdf to the fusion center and apply the PBDF at the fusion center. We have also carried out simulations for variance of w_n^k defined in equation (5.22) and shown in Fig. 5.3, together for all samples of n but with better visualization. We have done this

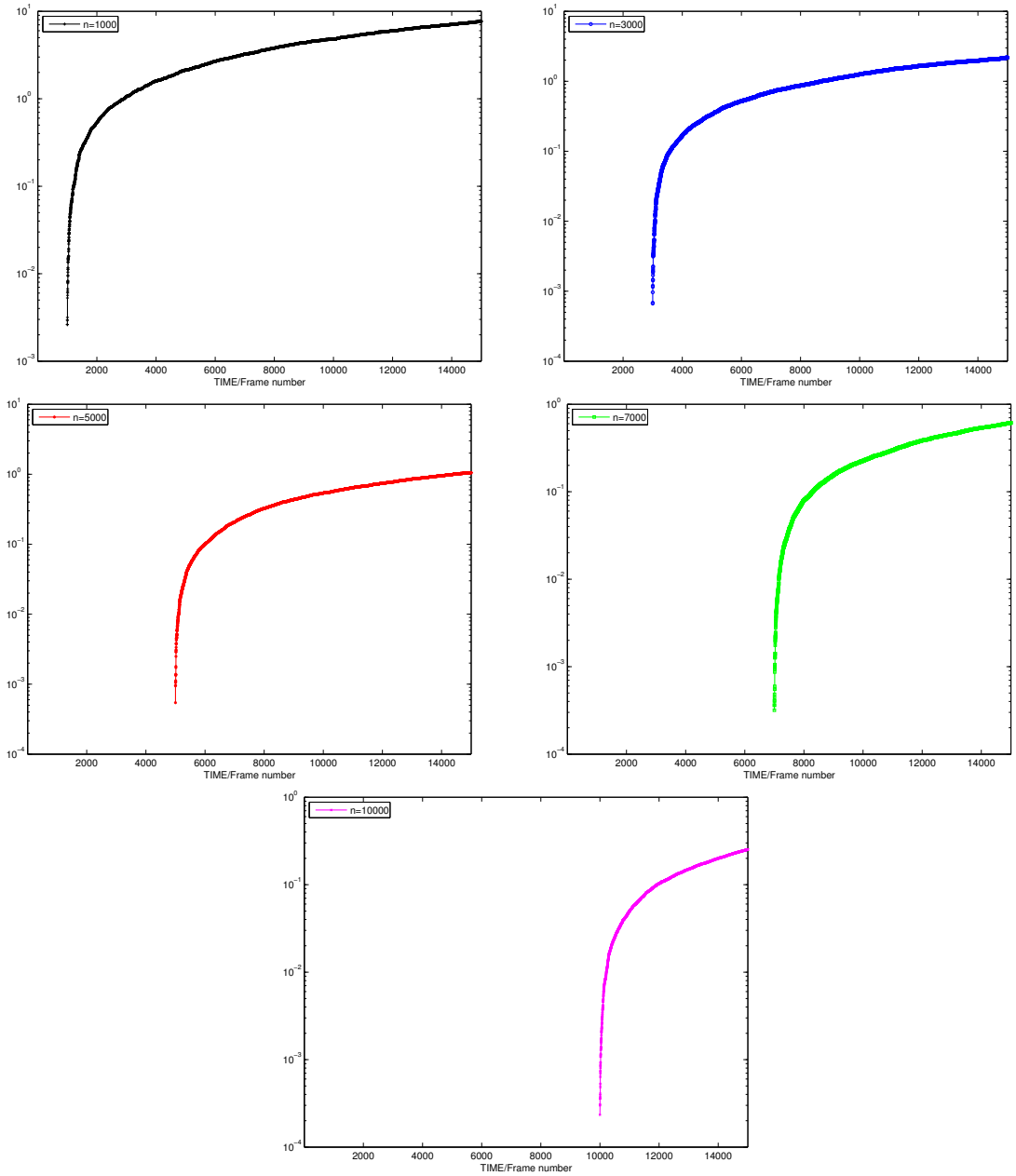

Figure 5.2: Simulation results of the individual variance of $\Theta_n^k = w_n^k X_n$ with $k = 15000$ and $n = 1000, 3000, 5000, 7000, 10000$. The vertical axis shows the logrithmic values of the individual variances.

for $k = 15000$ scans and n varying and shown on a semi log scale. We observe that when k is large enough the variance of w_n^k approaches a steady state value. We have plotted the simulation results of the terms T, $\sqrt{T_1(k)}$ and $T_2(k)$ involved in equation (5.22) in Fig.

5.4, and it is clear that these terms approach to one as required for convergence of the weights. We also include simulation results for variance of $w_n^k X_n$ with different values of n as shown in Fig. 5.5. It is clear that the variance of $w_n^k X_n$ depends on the variance of w_n^k. When the variance of w_n^k approaches zero, the variance of $w_n^k X_n$ approaches a steady state value and this is consistent in simulations if we put $w_n^k = 0$.

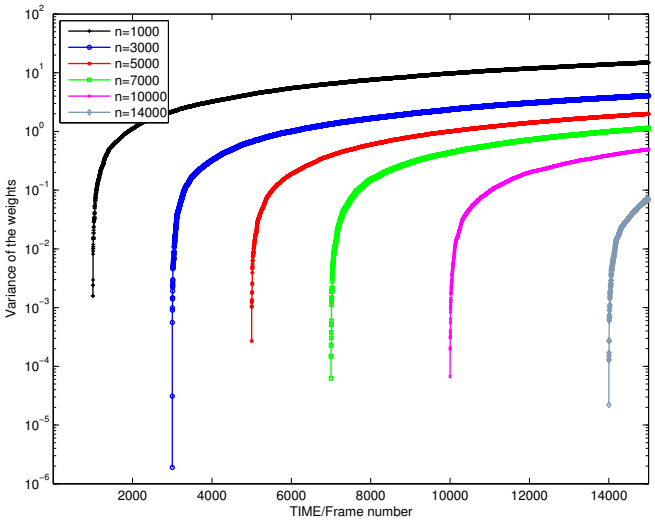

Figure 5.3: Convergence results of variance of weight w_n^k where $k = 15000$. The y-axis is on log scale while x-axis is on linear scale

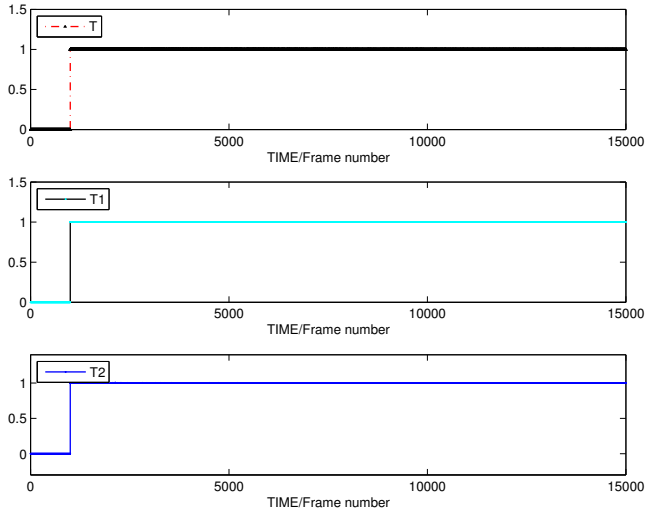

Figure 5.4: Convergence of T, $\sqrt{T_1(k)}$ and $T_2(k)$ to 1 for $n = 1000$

The root mean square errors for both filters are compared in Fig. 5.6 averaged over

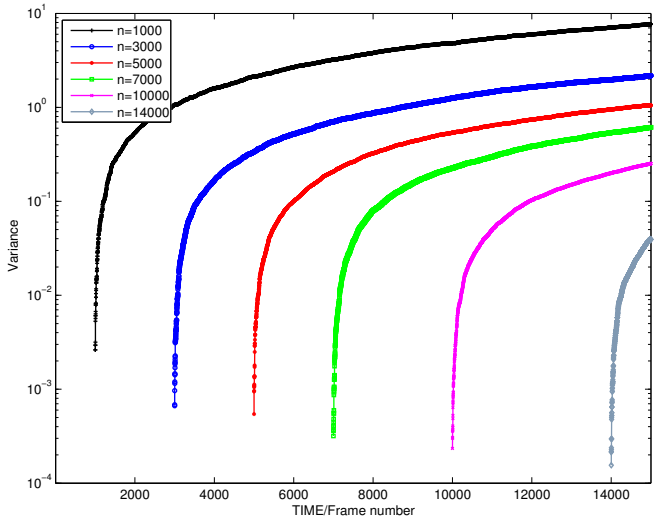

Figure 5.5: Convergence results of the variance of weight $w_n^k X_n$ where $k = 15000$. The y-axis is on log scale while x-axis is on linear scale

100 simulation runs. The time averaged RMSE values for local KF and the PBDF are 0.1296 and 0.1817 respectively. Since the loss of performance is small we can say PBDF is reliable in this problem. Figure 5.7 shows the asymptotic behavior of the variance of the weights of the samples drawn upto the current time. It is obvious that most recent samples have lower variances which is indication of the convergence. Similarly, variance of the product of weights and samples is shown in Fig. 5.8. By observing these results, we can say the variance of the PBDF converges.

Simulation results included in this section show that the PBDF has comparable performance with local optimal or centralized estimator and also convergence of the variance of the estimator as well as the weights. The mean of the proposed estimator is very close to the conditional mean of the optimal estimator.

5.6 Summary

In this chapter, we proposed a method of distributed tracking or estimation under communication constraints. In the proposed approach, we transmit only one particle or sample drawn from the local posterior pdf of state/parameter and transmit to the fusion center. At the fusion center, we calculate/update the weight of every received particle

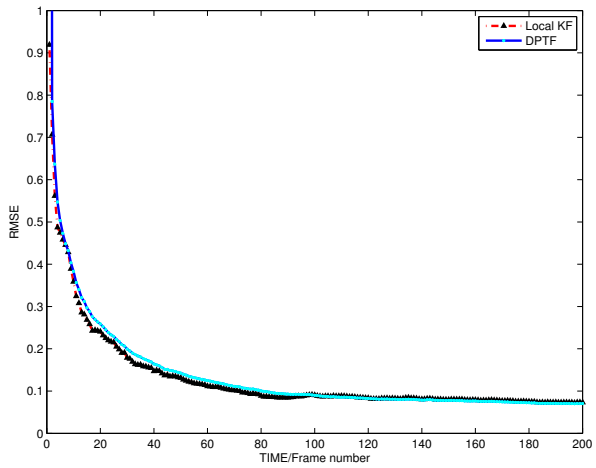

Figure 5.6: Performance comparison of local KF and proposed global Distributed particle type filter

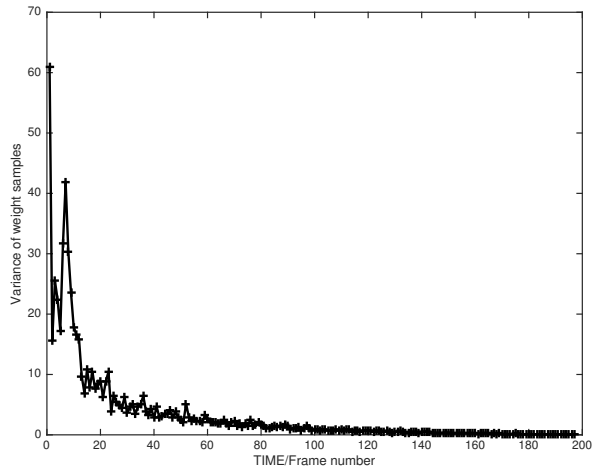

Figure 5.7: Asympotatic behaviour of the updated variances of the weights of the samples drawn upto the current time k

and our estimator is defined by a weighted sum of the samples. The calculation of the weights of samples is based on adaptive importance sampling techniques. At time k, we have a different importance density for all samples but the same target density. Since, we assume our samples are drawn independently but with non-identical distributions the conventional central limit theorem does not hold. Also convergence of the proposed estimator directly depends on convergence of the weights and is discussed in the next chapter. We have also included simulation results of variances of w_n^k and $w_n^k X_n$ and it can be seen that these variances show asymptotic convergence to some static value.

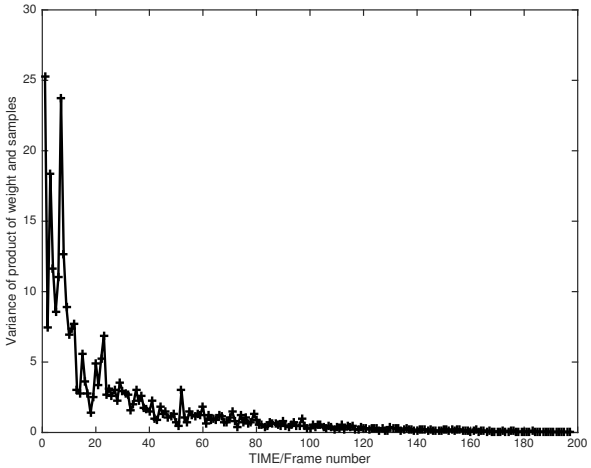

Figure 5.8: Asympotatic behaviour of the updated variances of the product of weights and samples drawn upto the current time k

We are working to extend this idea to the dynamic case where we assume moving targets. We hypothesize that samples received before the current time should follow target dynamics and then update their weights with current measurement information. Other possible future work is to use the proposed method for the multisensor case both for parameter estimation as well as for moving targets.

Chapter 6

Convergence of Particle Based Distributed Estimator

This chapter examines the performance of the particle based distributed estimator. Specifically, we derive convergence properties of the Particle Based Distributed Filter (PBDF) proposed in the previous chapter. We show that the mean and variance of the proposed estimator are bounded and these sequences converge. Since the variance of the estimator depends on the variance of the weights we study the variance of the weights as a first step. Also, we show that individual variance sequences of the product of weight and sample converge asymptotically.

6.1 Introduction

BOTH consistency and efficiency are considered critically important in the design of an estimator. The performance of the PBDF estimator is studied and compared with centralized estimator in the Chapter 5. Generally, consistency describes the behavior of an estimator for an infinite number of observations. For a non-random parameter x with true value x_0 and measurement set Y^k, an estimator $\hat{x}_k$ is consistent if the estimate of the parameter (which is a random variable) converges to the true value under some criteria [2]. The definition of a consistent estimator using a mean square criterion is given by

$$\lim_{k\to\infty} E\left[(\hat{x}_k - x_0)^2\right] = 0 \tag{6.1}$$

For a random parameter x, the definition of a consistent estimator slightly differs because we do not have true value and can be expressed as

$$\lim_{k\to\infty} E\left[(\hat{x}_k - x)^2\right] = 0 \tag{6.2}$$

where the expectation is taken over both the measurement set Y^k and the random quantity x.

In this chapter, we discuss the convergence of the estimator defined in equation (5.13). We prove that the first two moments of this estimator are asymptotically convergent. The convergence of the variance of the proposed estimator depends on the variance of the weights. We prove convergence of the variance of the weights Ψ_n^k and convergence of the variance of the product of the weights and samples Θ_n^k defined in equations (5.22) and (5.29).

The expected value of PBDF estimates converges to the local conditional mean estimate and from the basic Kalman filter consistency theory, the mean of the traditional Kalman filter converges. Therefore, the mean of PBDF converges. In this chapter we study the convergence of the variance of the proposed distributed filter. In the simulations discussed in the previous chapter, we observed that the variance of the proposed estimator appears to converge. We also observed that convergence of the variance was dependent on convergence of the variance of the weights. Moreover, simulation results showed that the variance of the weights of the samples Ψ_n^k converges for all instances of n. In this chapter, we prove convergence of variance of the weights and the PBDF.

Problem formulation and techniques to prove the convergence of the variance sequence is explained in section 2. In section 3, we include the results of the convergence of the variance sequence of the weights of the samples. Section 4 presents the convergence results of PBDF.

6.2 Problem Formulation

In order to prove convergence of the proposed estimator, we need to derive the convergence of its mean and variance. The proposed estimator with its associated parameters derived in the previous chapter are

$$\Delta_k = \frac{w_1^k X_1 + w_2^k X_2 + w_3^k X_3 + ... + w_k^k X_k}{k} \tag{6.3}$$

with the mean and variance of the weights w_n^k

$$\begin{aligned}\hat{\phi}_n^k &= 1 \\ \Psi_n^k &= \frac{1}{\sqrt[k-n]{4\pi R}} \frac{N(y_{n+1}; H\hat{x}_{n|n}, S_n - R/2)}{N^2(y_{n+1}; H\hat{x}_{n|n}, S_n)} \\ &\quad \times \prod_{i=n+2}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} - 1\end{aligned} \tag{6.4}$$

The expected value of the proposed estimator defined in equation (5.13) is given in equation (6.5) where the expectation has taken over all the samples. The expected value of the proposed estimator is equal to the local Kalman filter estimate as shown in equation (6.5).

$$E[\Delta_k] = \hat{x}_{k|k} \tag{6.5}$$

The variance of the PBDF is given as

$$\begin{aligned}Var[\Delta_k] &= \frac{1}{k^2} \sum_{n=1}^{k} \Theta_n^k \\ &= \frac{1}{k^2} \sum_{n=1}^{k} \frac{1}{(4\pi R)^{k-n/2}} \frac{N(y_{n+1}; H\hat{x}_{n|n}, S_n - R/2)}{N^2(y_{n+1}; H\hat{x}_{n|n}, S_n)} \\ &\quad \times \prod_{i=n+2}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2\end{aligned} \tag{6.6}$$

The parameters involved in above set of equations are rewritten as

$$\begin{aligned}\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - K_k H P_{k|k-1} \\ \bar{x}_{k|k} &= \bar{x}_{k|k-1} + \bar{K}_k(y_k - H\bar{x}_{k|k-1}) \\ \bar{P}_{k|k} &= \bar{P}_{k|k-1} - \bar{K}_k H \bar{P}_{k|k-1}\end{aligned}$$

where $K_k = P_{k|k-1}H'S_{k-1}^{-1}$ and $\bar{K}_k = \bar{P}_{k|k-1}H'\bar{S}_{k-1}^{-1}$ are the gains for different measurement noise variances.

By observing equation (6.5), the mean of the proposed estimator is equal to the local estimator conditional mean which is the result of the standard Kalman filter equation.

The unconditional mean of the parameter is convergent to the true value when integrated over all measurements [2] as shown in Appendix A.4. We need to prove that the variance of the proposed estimator is bounded and finite for some large value of k. Firstly, we derive expressions for Θ_n^k in terms of Ψ_n^k. Secondly, we study the convergence properties of the sequence Θ_n^k for all values of n and large k. The expression for Θ_n^k in terms of the variance of the weights in equation (6.7) is (Appendix B.1),

$$\Theta_n^k = \left[\Psi_n^k + 1\right] \times \left[\bar{P}_{k|k} + \bar{x}_{k|k}^2\right] - \hat{x}_{k|k}^2 \tag{6.7}$$

Equation (6.7) shows the significance of weight variance for the calculation of variance of the proposed estimator. We first show that the sequences Θ_n^k and Ψ_n^k are Cauchy Sequences (CS). We start discussing the properties of the Ψ_n^k sequence and prove it as a CS and then we prove that Θ_n^k is also a CS. By observing equation (6.7), it is clear that Θ_n^k is convergent sequence if Ψ_n^k is convergent.

6.3 Convergence of the Variance of the Weights

In this section, we discuss convergence properties of the variance sequence of sample weights w_n^k. Our main focus is to show that the sequence of the variance of the weights Ψ_n^k is convergent for every $n = 1, 2, 3, ..k$ where $k \geqslant n$ and n is large positive integer as defined earlier. In this section, we prove that the sequences Ψ_n^k are bounded and convergent.

The variance of the weights w_n^k can be re-written as

$$\begin{aligned}
\Psi_n^k &= \sqrt{\frac{S_{k-1}}{R}} \frac{N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{k-1|k-1}H' + R/2)}{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1|k-1}/2)} \left[\Psi_n^{k-1} + 1\right] - 1 \\
\Psi_n^k &= T_k \left[\Psi_n^{k-1} + 1\right] - 1 \\
\Psi_n^k &= T_k \left[\Psi_n^{k-1} + 1\right] - 1 \\
\Psi_n^k &= T_k \Psi_n^{k-1} + T_k - 1
\end{aligned} \tag{6.8}$$

where

$$T_k = \sqrt{\frac{S_{k-1}}{R}} \frac{N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{k-1|k-1}H' + R/2)}{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1|k-1}/2)} \tag{6.9}$$

The above recursion can be written in terms of Ψ_n^0 by backward substitution as follows

$$\begin{aligned}
\Psi_n^k &= T_k\Psi_n^{k-1} + T_k - 1 \\
\Psi_n^k &= T_k[T_{k-1}\Psi_n^{k-2} + T_{k-1} - 1] + T_k - 1 \\
\Psi_n^k &= T_kT_{k-1}\Psi_n^{k-2} + T_kT_{k-1} - T_k + T_k - 1 \\
\Psi_n^k &= T_kT_{k-1}\Psi_n^{k-2} + T_kT_{k-1} - 1
\end{aligned} \tag{6.10}$$

Now substituting the recursive form of Ψ_n^{k-2}, we obtain

$$\begin{aligned}
\Psi_n^k &= T_kT_{k-1}\Psi_n^{k-2} + T_kT_{k-1} - 1 \\
\Psi_n^k &= T_kT_{k-1}[T_{k-2}\Psi_n^{k-3} + T_{k-2} - 1] + T_kT_{k-1} - 1 \\
\Psi_n^k &= T_kT_{k-1}T_{k-2}\Psi_n^{k-3} + T_kT_{k-1}T_{k-2} - T_kT_{k-1} + T_kT_{k-1} - 1 \\
\Psi_n^k &= T_kT_{k-1}T_{k-2}\Psi_n^{k-3} + T_kT_{k-1}T_{k-2} - 1
\end{aligned} \tag{6.11}$$

Similarly, substituting Ψ_n^{k-3}, Ψ_n^{k-4} to Ψ_n^1, we obtain

$$\begin{aligned}
\Psi_n^k &= \prod_{i=1}^{k} T_i\Psi_n^0 + \prod_{i=1}^{k} T_i - 1 \\
&= \prod_{i=1}^{k} T_i(\Psi_n^0 + 1) - 1
\end{aligned} \tag{6.12}$$

In order to prove convergence of the sequence of Ψ_n^k over time, we examine the asymptotic behavior of the product term $\prod_{i=1}^{k} T_i$.

Lemma 6.1. *Let $T_k \neq 0$, for all $k = 1, 2, 3, ...$ then $\prod_{k=1}^{\infty} T_k$ converges to a nonzero limit iff the series $a_k = \sum_{k=1}^{\infty} log(T_k)$ converges.*

Proof. Proof is given in Chapter 6 [57]. □

6.3.1 The Series a_k is Convergent

In this subsection, we study the convergence property of the series $a_k = \sum_{k=1}^{\infty} log(T_k)$ which is sequence of the sums. Let $b_k = \log(T_k)$, then $a_k = \sum_{k=1}^{\infty} b_k$.

Suppose, we have a series $a_k = \sum_{i=1}^{k} log(T_k)$ and we wish to prove convergence of the partial sums a_k. The series a_k converges if and only if the partial sums $a_1, a_2, a_3, ..., a_k$ converge. Also for all instances of k, the series a_k is convergent series if the Cauchy property holds, i.e., for every $m, n \geqslant N$, $|a_m - a_n| \leqslant \epsilon$. As a first step, we find the logarithm of T_k defined in equation (6.9) $b_k = \log(T_k)$. Hence

$$\begin{aligned} T_k &= \sqrt{\frac{S_{k-1}}{R}} \frac{N(y_k; H\bar{x}_{k-1|k-1}, \bar{S}_{k-1})}{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1}/2)} \\ b_k = \ log(T_k) &= \log\left(\sqrt{\frac{S_{k-1}}{R}} \frac{N(y_k; H\bar{x}_{k-1|k-1}, \bar{S}_{k-1})}{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1}/2)}\right) \end{aligned} \tag{6.13}$$

Thus

$$\begin{aligned} b_k &= \frac{1}{2}\Big\{ \log(S_{k-1}) - \log(R) - \log(\bar{S}_{k-1}) + \log(S_k/2) - \frac{(y_k - H\bar{x}_{k-1|k-1})^2}{\bar{S}_{k-1}} \\ &\quad + \frac{(y_k - H\hat{x}_{k-1|k-1})^2}{S_{k-1}/2}\Big\} \\ &= \frac{1}{2}\Big\{2\log(S_{k-1}) - \log(R) - \log(\bar{S}_{k-1}) - \log(2) - \frac{\left(y_k - H\bar{x}_{k-1|k-1}\right)^2}{\bar{S}_{k-1}} \\ &\quad + \frac{2(y_k - H\hat{x}_{k-1|k-1})^2}{S_{k-1}}\Big\} \\ &= \frac{1}{2}\left[J_k - \bar{Q}_k + \hat{Q}_k\right] \end{aligned} \tag{6.14}$$

where $J_k = 2\log(S_{k-1}) - \log(R) - \log(\bar{S}_{k-1}) - \log(2)$, $\bar{Q}_k = \left(y_k - H\bar{x}_{k-1|k-1}\right)^2 / \bar{S}_{k-1}$ and $\hat{Q}_k = 2(y_k - H\hat{x}_{k-1|k-1})^2 / S_{k-1}$

We can further simplify the above expressions for J_k, $\bar{Q}_k$ and $\hat{Q}_k$ by substituting $\bar{S}_{k-1}$, S_{k-1}, $\bar{x}_{k-1|k-1}$ and $\hat{x}_{k-1|k-1}$ as derived in Appendix B.2,

$$\begin{aligned} J_{k+1} &= 2\log(S_k) - \log(R) - \log(\bar{S}_k) + \log(2) \\ &= 2\log\left(\frac{kP_oR + R^2}{(k-1)P_0 + R}\right) - \log(R) - \log\left(\frac{2kP_oR + R^2}{2[2(k-1)P_0 + R]}\right) - \log(2) \\ &= 2\log(kP_oR + R^2) - 2\log((k-1)P_0 + R) - \log(R) - \log(2kP_oR + R^2) \\ &+ \log(2[2(k-1)P_0 + R]) - \log(2) \end{aligned} \tag{6.15}$$

Further simplifying we obtain

$$\begin{aligned} J_{k+1} &= 2\log(kP_o + R) + 2\log(R) - 2\log((k-1)P_0 + R) - \log(R) - \log(2kP_o + R) \\ &- \log(R) + \log(2(k-1)P_0 + R) + \log(2) - \log(2) \\ &= 2\log(kP_o + R) - 2\log((k-1)P_0 + R) - \log(2kP_0 + R) + \log(2(k-1)P_0 + R) \\ &= 2\log\left(\frac{kP_o + R}{(k-1)P_0 + R}\right) - \log\left(\frac{2kP_0 + R}{2(k-1)P_0 + R}\right) \\ &= 2\hat{J}_{k+1} - \bar{J}_{k+1} \end{aligned}$$

where $\hat{J}_{k+1} = \log(kP_o + R/(k-1)P_0 + R)$ and $\hat{J}_{k+1} = \log(2kP_0 + R/2(k-1)P_0 + R)$

$$\begin{aligned} \bar{Q}_{k+1} &= \frac{(y_k - \bar{x}_{k|k})^2}{\bar{S}_k} \\ &= \frac{2[2(k-1)P_0 + R]}{2kP_0R + R^2}(y_k - \bar{x}_{k|k})^2 \\ &= \frac{2[2kP_0 - 2P_0 + R]}{(2kP_0 + R)R}(y_k - \bar{x}_{k|k})^2 \end{aligned} \tag{6.16}$$

Multiplying and dividing by k.

$$\bar{Q}_{k+1} = \frac{2[2P_0 - 2P_0/k + R/k]}{(2P_0 + R/k)R}(y_k - \bar{x}_{k|k})^2$$

For very large k, the fractions $2P_0/k$ and R/k approach zero, thus

$$\bar{Q}_{k+1} \to \frac{2}{R}(y_k - \bar{x}_{k|k})^2 \tag{6.17}$$

substituting S_k in the formula of $\hat{Q}_{k+1}$ we obtain

$$\begin{aligned}\hat{Q}_{k+1} &= \frac{2(y_k - \hat{x}_{k|k})^2}{S_k} \\ &= \frac{(k-1)P_0 + R}{kP_0R + R^2} 2(y_k - \hat{x}_{k|k})^2 \\ &= \frac{kP_0 - P_0 + R]}{(kP_0 + R)R} 2(y_k - \hat{x}_{k|k})^2\end{aligned}$$

Multiplying and dividing by k.

$$\hat{Q}_{k+1} = \frac{P_0 - P_0/k + R/k}{(P_0 + R/k)R} 2(y_k - \hat{x}_{k|k})^2$$

For very large k, the fractions P_0/k and R/k approach zero, thus

$$\hat{Q}_{k+1} \to \frac{2}{R}(y_k - \hat{x}_{k|k})^2 \tag{6.18}$$

Lemma 6.2. $\lim_{k\to\infty} |\hat{x}_{k|k} - \bar{x}_{k|k}| = 0$

Proof. The expressions for $\bar{x}_{k|k}$ and $\hat{x}_{k|k}$ from Appendix (B.2) are

$$\begin{aligned}\bar{x}_{k|k} &= \frac{R}{2kP_0 + R} x_0 + \frac{2P_0}{2kP_0 + R} \sum_{i=1}^{k} y_i \\ \hat{x}_{k|k} &= \frac{R}{kP_0 + R} x_0 + \frac{P_0}{kP_0 + R} \sum_{i=1}^{k} y_i\end{aligned}$$

$2kP_0 >> R$ and $kP_0 >> R$ hence ignoring R from denominators of the above expressions.

$$\begin{aligned}\bar{x}_{k|k} &\simeq \frac{R}{2kP_0} x_0 + \frac{P_0}{kP_0} \sum_{i=1}^{k} y_i \\ \hat{x}_{k|k} &\simeq \frac{R}{kP_0} x_0 + \frac{P_0}{kP_0} \sum_{i=1}^{k} y_i\end{aligned}$$

$R/2kP_0 \approx R/kP_0$ hence $\bar{x}_{k|k} - \hat{x}_{k|k} \simeq \frac{Rx_0}{2P_0}\frac{1}{k}$ thus

$$\lim_{k\to\infty} |\hat{x}_{k|k} - \bar{x}_{k|k}| = 0 \tag{6.19}$$

□

Lemma 6.3. $\lim_{m\to\infty} |\hat{Q}_{m+1} - \bar{Q}_{m+1}| \to 0$

Proof. Using the the results of (6.18) and (6.17).

$$\begin{aligned} \bar{Q}_{k+1} &\to \frac{2}{R}(y_k - \bar{x}_{k|k})^2 \\ &\to \frac{2}{R}\left(y_k^2 + \bar{x}_{k|k}^2 - 2y_k\bar{x}_{k|k}\right) \end{aligned} \tag{6.20}$$

Substituting the expression of $\bar{x}_{k|k}$ in above equation, we obtain

$$\begin{aligned} \bar{Q}_{k+1} &\to \frac{2}{R}\left[y_k^2 + \left(\frac{R}{2kP_0}x_0 + \frac{P_0}{kP_0}\sum_{i=1}^{k} y_i\right)^2 - 2y_k\left(\frac{R}{2kP_0}x_0 + \frac{P_0}{kP_0}\sum_{i=1}^{k} y_i\right)\right] \\ &\to \frac{2}{R}\left[y_k^2 + \bar{\epsilon}_k\right] \\ &\to \frac{2}{R}\left[\mu + \bar{\epsilon}_k\right] \end{aligned} \tag{6.21}$$

Similarly, we can write the expression for $\hat{Q}_{k+1}$

$$\begin{aligned} \hat{Q}_{k+1} &\to \frac{2}{R}(y_k - \hat{x}_{k|k})^2 \\ &\to \frac{2}{R}\left(y_k^2 + \hat{x}_{k|k}^2 - 2y_k\hat{x}_{k|k}\right) \end{aligned} \tag{6.22}$$

Substituting the expression of $\hat{x}_{k|k}$ in above equation, we obtain

$$\begin{aligned} \hat{Q}_{k+1} &\to \frac{2}{R}\left[y_k^2 + \left(\frac{R}{kP_0}x_0 + \frac{P_0}{kP_0}\sum_{i=1}^{k} y_i\right)^2 - 2y_k\left(\frac{R}{2kP_0}x_0 + \frac{P_0}{kP_0}\sum_{i=1}^{k} y_i\right)\right] \\ &\to \frac{2}{R}\left[y_k^2 + \hat{\epsilon}_k\right] \\ &\to \frac{2}{R}\left[\mu + \hat{\epsilon}_k\right] \end{aligned} \tag{6.23}$$

For a large k, both $\hat{\epsilon}_k$ and $\hat{\epsilon}_k$ approach 0, hence

$$\begin{aligned} \bar{Q}_{k+1} &\to \frac{2}{R}\mu \\ \hat{Q}_{k+1} &\to \frac{2}{R}\mu \end{aligned} \tag{6.24}$$

Based on the above working, expression for $\left|\hat{Q}_{k+1} - \bar{Q}_{k+1}\right|$ becomes

$$\begin{aligned}\left|\hat{Q}_{k+1} - \bar{Q}_{k+1}\right| &\to \frac{2}{|R|}\left|\mu + \hat{\epsilon}_k - \mu - \bar{\epsilon}_k\right| \\ &\to \frac{2}{|R|}\left|\hat{\epsilon}_k - \bar{\epsilon}_k\right|\end{aligned}$$

By applying limits, we can write

$$\lim_{m\to\infty}\left|\hat{Q}_{k+1} - \bar{Q}_{k+1}\right| \to 0 \tag{6.25}$$

□

Lemma 6.4. $\lim_{m,n\to\infty}|L(m,n) - M(m,n)| \to 0$

Proof. We can write the expressions for $L(m,n)$ using reduced forms of $\sum_{i=m+1}^{n}\hat{J}_i$ and $\sum_{i=m+1}^{n}\hat{J}_i$ derived in Appendix B.3.

$$\begin{aligned}L(m,n) &= \sum_{i=m+1}^{n}\hat{J}_i - 1/2\sum_{i=m+1}^{n}\bar{J}_i \\ &= \log\left(\frac{nP_o + R}{mP_0 + R}\right) - 1/2\log\left(\frac{2nP_o + R}{2mP_0 + R}\right) \\ &= \log\left(\frac{nP_o + R}{mP_0 + R}\right) - \log\left(\sqrt{\frac{2nP_o + R}{2mP_0 + R}}\right) \\ &= \log\left(\frac{nP_o + R}{mP_0 + R}\sqrt{\frac{2mP_o + R}{2nP_0 + R}}\right)\end{aligned}$$

Using the argument, for very large values of m and n, we can say that $nP_0 >> R$ and

$mP_0 >> R$, therefore we can ignore R and simplify the expressions as

$$\begin{aligned} L(m,n) &= \sum_{i=m+1}^{n} \hat{J}_i - 1/2 \sum_{i=m+1}^{n} \bar{J}_i \\ &= \log\left(\frac{nP_o}{mP_0}\sqrt{\frac{2mP_0}{2nP_0}}\right) \\ &= \log\left(\sqrt{\frac{n}{m}}\right) \\ &= 1/2\log\left(\frac{n}{m}\right) \end{aligned} \tag{6.26}$$

As we know that $m >> 0$ and $n >> 0$, therefore $\log(n/m) \to 0$. Using this argument, the term $L(m,n)$ defined above approaches to 0.

$$\begin{aligned} L(m,n) &= \sum_{i=m+1}^{n} \hat{J}_i - 1/2 \sum_{i=m+1}^{n} \bar{J}_i \\ &\to 0 \end{aligned} \tag{6.27}$$

From equation (6.31),

$$\begin{aligned} 2M(m,n) &= \sum_{i=m+1}^{n} \left(\hat{Q}_i - \bar{Q}_i\right) \\ &= \left(\hat{Q}_{m+1} - \bar{Q}_{m+1}\right) + \left(\hat{Q}_{m+2} - \bar{Q}_{m+2}\right) + ...+ \\ &\left(\hat{Q}_{n-1} - \bar{Q}_{n-1}\right) + \left(\hat{Q}_n - \bar{Q}_n\right) \end{aligned} \tag{6.28}$$

Using the lemma 6.3, we can rewrite the limiting value of $2M$ as

$$\begin{aligned} 2M(m,n) &= \sum_{i=m+1}^{n} \left(\hat{Q}_i - \bar{Q}_i\right) \\ 2M(m,n) &\to 0 \\ M(m,n) &\to 0 \end{aligned} \tag{6.29}$$

Using the convergence of results of $L(m,n)$ and $M(m,n)$ in equations (6.27) and (6.29)

we can write the equation in limiting case as

$$\lim |L(m,n) + M(m,n)| \to 0 \tag{6.30}$$

□

Theorem 6.1. *For $k \geqslant N$ the sequence a_k is a Cauchy Sequence where $a_k = \sum_{i=1}^{k} b_i$ and $\prod_{i=1}^{k} T_i$ converges.*

Proof. We can write $a_k = \sum_{i=1}^{k} b_i$. For $m, n \geqslant N$ and $n > m$, we can write

$$\begin{aligned}
|a_m - a_n| &= |\sum_{i=1}^{m} b_i - \sum_{i=1}^{n} b_i| \\
&= |\sum_{i=m+1}^{n} b_i| \\
&= |\frac{1}{2} \sum_{i=m+1}^{n} \left(J_i - \bar{Q}_i + \hat{Q}_i\right)| \\
&= |\frac{1}{2} \sum_{i=m+1}^{n} J_i + \frac{1}{2} \sum_{i=m+1}^{n} \left(\hat{Q}_i - \bar{Q}_i\right)| \\
&= |\frac{1}{2}(\sum_{i=m+1}^{n} \left[2\hat{J}_i - \bar{J}_i\right] + \frac{1}{2} \sum_{i=m+1}^{n} \left(\hat{Q}_i - \bar{Q}_i\right)| \\
&= |\sum_{i=m+1}^{n} \hat{J}_i - \frac{1}{2} \sum_{i=m+1}^{n} \bar{J}_i + \frac{1}{2} \sum_{i=m+1}^{n} \left(\hat{Q}_i - \bar{Q}_i\right)| \\
&= |L(m,n) + M(m,n)|
\end{aligned} \tag{6.31}$$

where $L(m,n) = \sum_{i=m+1}^{n} \hat{J}_i - \frac{1}{2} \sum_{i=m+1}^{n} \bar{J}_i$ and $M(m,n) = \frac{1}{2} \sum_{i=m+1}^{n} \left(\hat{Q}_i - \bar{Q}_i\right)$.

Using the results of lemma 6.4, it is clear that of $|L(m,n) + M(m,n)| \to 0$. Therefore, it is proven that the partial sum sequence a_k is a Cauchy sequence.

We proved the convergence of partial sums of a_k based on the fact that a_k is Cauchy Sequence and any CS is an a convergent sequence. Using lemma 6.1, it is evident that the product term $\prod_{i=1}^{k} T_i$ is convergent. □

Theorem 6.2. Convergence of the variance of the weights *For $k \geqslant N$, the sequence Ψ_n^k is a convergent sequence for every $n = 1, 2, 3, ..., k$. where $\Psi_n^k = var(w_n^k | Y^k)$ is the variance of the*

weights.

Proof. We recall the expression of Ψ_n^k as function of initial conditions and re-write the equation below:

$$\begin{aligned}\Psi_n^k &= \prod_{i=1}^{k} T_i \Psi_n^0 + \prod_{i=1}^{k} T_i - 1 \\ &= \mathrm{Y} \times \left(\Psi_n^0 + 1\right) - 1\end{aligned} \tag{6.32}$$

By observing the recursive equation of Ψ_n^k in equation (6.32), if Y converges then the sequence Ψ_n^k also converges since Ψ_n^0 and 1 are constant. Using theorem 6.1, the sequence of the product term $\mathrm{Y} = \prod_{i=1}^{k} T_i$ is convergent sequence. Hence the variance of the weight sequences Ψ_n^k are convergent for all n. □

6.4 Convergence of the Variance of the PBDF

In order to prove the convergence of the variance of proposed estimator defined in equation (5.13), we need to examine the convergence of individual variances of the random variables G_i which are the product of the weights and samples.

$$\begin{aligned}\Delta_k &= \frac{w_1^k X_1 + w_2^k X_2 + w_3^k X_3 + ... + w_k^k X_k}{k} \\ &= \frac{G_1^k + G_2^k + G_3^k + ... + G_k^k}{k}\end{aligned} \tag{6.33}$$

As defined earlier, the variance of the proposed estimator is

$$\begin{aligned}Var[\Delta_k] &= \frac{1}{k^2}\left[Var(G_1^k) + Var(G_2^k) + Var(G_3^k) + ... + Var(G_k^k)\right] \\ &= \frac{1}{k^2}\left[\Theta_1^k + \Theta_2^k + \Theta_3^k + ... + \Theta_{k-1}^k + \Theta_k^k\right] \\ &= \frac{1}{k^2}\sum_{i=1}^{k}\Theta_i^k\end{aligned} \tag{6.34}$$

Theorem 6.3. Convergence of variance of G_n^k: *For $k \geqslant N$, the sequence of variances of the*

random variables G_n^k denoted by Θ_n^k is convergent sequence where $\Theta_n^k = Var(w_n^k X_n | Y^k)$ for $n = 1, 2, 3..., k$. This implies the convergence of $Var(\Delta_k)$.

Proof. The proof of this theorem is based on the convergence of variance of the weights. Consider the sequence

$$\Theta_n^k = \left[\Psi_n^k + 1\right] \times \left[\bar{P}_{k|k} + \bar{x}_{k|k}^2\right] - \hat{x}_{k|k}^2 \tag{6.35}$$

The sequence Θ_n^k is convergent if the terms Ψ_n^k, $\bar{P}_{k|k}$, $\bar{x}_{k|k}^2$ and $\hat{x}_{k|k}^2$ approach to steady state values. We discuss the convergence of each term. The first term Ψ_n^k is the variance of weights and plays vital role in the convergence of estimator variance. The convergence of Ψ_n^k is already proved in Theorem 6.2 discussed in previous section. The other parameters $\bar{P}_{k|k}$, $\bar{x}_{k|k}^2$ and $\hat{x}_{k|k}^2$ are the variance and estimates resulted from a Kalman filter and consistency of the Kalman filter is well studied in the literature [2].

To conclude, all the four terms involved in expression of Θ_n^k are convergent and hence the sequence itself is convergent. This result leads to convergence of variance of the proposed estimator. □

6.5 Summary

In this chapter, we discussed the convergence of the PBDF. We show that the asymptotic value of the mean of the weights is 1 while the asymptotic value of the conditional mean of the proposed estimator is equal to the optimal conditional mean of the local estimator. We have also seen that the unconditional mean approaches the true value of the parameter. We derived equations for the variance of the proposed estimator in terms of the variance of the weights. We proved that the sequence of the variance of the weight is a Cauchy sequence. This result leads us to the conclusion that the sequence of the variances of the proposed estimator is also convergent.

Chapter 7
Conclusion

This chapter concludes the thesis and also includes discussion on the possible future work.

IN this thesis, distributed estimation or tracking under communication constraints was explored. We considered tracking both in cluttered and non-cluttered surveillance environments. We made three major contributions

- Distributed tracking using scaled equivalent measurements
- Distributed tracking using compressed measurements in cluttered environments
- Distributed tracking using particle transmission to the fusion center

We considered a single sensor node capable of performing local tracking which is connected to the fusion center. The communication link between the sensor node and the fusion center is assumed to be ideal, i.e., there is no loss of information. However, communication channels come with data rate limitations. We explored various distributed tracking algorithms by studying the effect of minimal measurement information across these communication links to the fusion center. We studied the trade-off between the available bandwidth and the tracking performance. The objective of this study was to analyze the communication constraints in the distributed architecture of a target tracking problem.

We commenced by introducing the problem in Chapter 1. A review of target tracking methods was covered in Chapter 2. In Chapter 3, we discussed the transmission of equivalent measurements (tracklets) as measurement information in a distributed tracking architecture when using low data rate communication. We observed that sending scaled equivalent measurements is similar to the transmission of the full state equivalent measurements in terms of tracking performance. The bandwidth required to transmit

the scaled equivalent measurements per scan is almost half of that required to transmit the full state equivalent measurements. Therefore, we can send scaled equivalent measurements more often using the same bandwidth as compared to the full state equivalent measurement which can be advantageous under certain circumstances. We also studied the encoding effect on the tracking performance. We observed that it is relatively better to send less often with a higher number of encoding bits rather than sending more often with a lower number of encoding bits. We proposed that sending information with an N-step scan interval rather than every scan is a good choice for optimizing available bandwidth. By doing so, we distributed the bandwidth to encode the scaled equivalent measurements based on the frequency of transmission. We discussed the idea of transmission of scaled equivalent measurements with a given budget of encoding bits for transmission. We compared the tracking performance of N step transmission with smaller and larger N values while keeping the same bandwidth budget for a fixed time of tracking a target. Future work in this direction could incorporate an adaptive N based on the information contribution of the equivalent measurements.

In the Chapter 4, we studied distributed tracking in the presence of a clutter environment. The objective was again maintaining reasonable tracking performance in the presence of constraints on the channel bandwidth. The structure of the tracking system was similar to the one used in Chapter 3 except we dealt with clutter measurements. We proposed the transmission of a weighted sum of clutter and target measurements to the fusion center. By doing so, we only need to transmit one weighted sum containing both clutter and target measurement information. The scaling factor of each measurement (target or clutter) comes from the likelihood of standard Probabilistic Data Association filter. We used the standard PDA filter at the sensor node and proposed a recursive Bayesian Compressed Filter at the fusion center. The proposed Compressed Filter uses the compressed measurement and extracts clutter as well as target information to update the global track. The Compressed Filter is a recursive filter and its implementation is based on Monte Carlo integration. In fact, we compute various intractable integrals using the Monte Carlo integration and by substituting clutter measurement integrals with new random variables resulting from the sum of clutter measurements. The implemen-

tation of this Compressed Filter is carried out using a Gaussian mixture. We commenced with one prior Gaussian and the number of Gaussians increases with the time . The estimated value of the state was the weighted sum of Gaussians. We retained only significant weighted Gaussians by discarding the low weighted Gaussians. We compared our proposed Compressed Filter with the local PDA filter and performance was comparable to the sub-optimal PDA. Keeping in mind that we transmit only one weighted measurement to the fusion center and the loss of the performance in presence of communication constraint was negligible.

The preliminary work in the Chapter 4 was only implemented for one dimensional tracking in which target dynamics were simple, i.e., one target was considered moving on a straight line with constant velocity. As a future work, this can be extended to tracking in higher dimensions. It would be interesting to look at the performance of our proposed filter with target maneuvers. Also, this work can be extended to take into account the probability of target existence and multiple targets.

The third major contribution of this thesis was presented in the Chapter 5. We proposed a novel idea of distributed tracking based on transmission of particles to the fusion center. At the fusion center, we proposed a particle based distributed filter which uses the received particles along with previously received particles to estimate the parameter of interest. Our proposed PBDF does not depend on the type of filter used at the sensor node. The only information it needs is samples (particles) of the local posterior probability distribution function. The proposed filter updates the weights of all received samples and calculates the weight for a recently received sample. The estimated value of the parameter is then the weighted sum of all received samples. The calculation of the weights is a challenging step because as time progresses, every sample has a different sampling or proposal distribution and a changing target distribution. Therefore, the conventional importance sampling techniques cannot be used. We used adaptive importance sampling techniques to calculate the weights of the samples and derived expressions in recursive forms so that computations can be reduced. In simulations, we compared the tracking performance with local/optimal standard Kalman filter and results are promising for the simplest case of parameter estimation. To make confident comments about our proposed

method, we studied the convergence of the estimator using simulations in Chapter 5. The expressions of the variance of the proposed methods strongly suggested its dependence on the weight statistics. Using simulations and intuitive arguments, we proved that our estimator is stable.

Study of the consistency of the PBDF was the topic of Chapter 6. We proved that the mean and variance of our proposed PBDF were bounded and hence converged to some steady state values. Based on the expressions of the mean and variances of the proposed estimator, we started work on the convergence of the proposed PBDF. The mean value of the proposed estimator asymptotically matches with local conditional mean while variance depends on the variance of the weights. Firstly, we proved the convergence of the variance of the weights and then proved the convergence of the proposed PBDF. We used knowledge of the Kalman filter consistency to prove the results in Chapter 6. Simulation results were included in the Chapter 5 were consistent with the mathematical derived results in Chapter 6.

The distributed estimation/filtering based on the particle transmission can be extended in many directions. An important extension to this work is for the dynamical case of target tracking. A second extension to the topic can be studying this approach for the multi sensor case. Also, we can study the trad-off between transmitting more than one particles at a time and the tracking performance. We have not studied the rate of convergence and of course this is an important future topic.

Appendix A

Derivations for the PBDF

A.1 Important Formulae List

The product of two Gaussian distributions with the same parameters is given by :

$$N^2(y_2; HX_1, R) = \frac{1}{\sqrt{4\pi R}} N\left(y_2; HX_1, R/2\right) \tag{A.1}$$

The division of two Gaussian densities having same mean and different variance is given by

$$\begin{aligned} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N(y_2; H\hat{x}_{1|1}, S_1)} &= \sqrt{\frac{4\pi S_1^2}{R}} N\left(y_2; H\hat{x}_{1|1}, 2\frac{(S_1 - R/2)S_1}{R}\right) \\ &= N(y_2; H\hat{x}_{1|1}, 2\tilde{S_1}S_1/R) \end{aligned} \tag{A.2}$$

where $\tilde{S_1} = S_1 - R/2$

A.2 The Statistic of Weights (Section 5.4.1)

In this appendix, we derive the variances of w_n^k for each $n = 1, 2, 3, ..., n$. The variance of w_1^1 is zero as this weight is 1, i.e., a constant. At $k = 2$: We have two samples with weights w_1^2 and w_2^2. The variance of w_2^2 is zero. The conditional variance of $[w_1^2|Y^2]$ is derived as

below.

$$
\begin{aligned}
Var(w_1^2|Y^2) &= E\left[(w_1^2|Y^2)^2\right] - (E[w_1^2|Y^2])^2 \\
&= \int \left(w_1^2\right)^2 N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - 1 \\
&= \int \frac{N^2(y_2; HX_1, R)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - 1 \\
&= \frac{1}{\sqrt{4\pi R}} \int \frac{N(y_2; HX_1, R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - 1 \\
&= \frac{1}{\sqrt{4\pi R}} \int \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \bar{x}_{2|2}, \bar{P}_{2|2}) dX_1 - 1 \\
&= \frac{1}{\sqrt{4\pi R}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} - 1
\end{aligned}
\tag{A.3}
$$

The unconditional variance of w_1^2 is

$$
\begin{aligned}
E[(w_1^2)^2] &= \int\int\int (w_1^2)^2 N(X_1; \hat{x}_{1|1}, P_{1|1}) N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1 \\
&= \int\int\int \frac{N^2(y_2; HX_1, R)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{1|1}, P_{1|1}) N(y_2; H\hat{x}_{1|1}, S_1) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1 \\
&= \int\int\int \frac{N^2(y_2; HX_1, R)}{N(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{1|1}, P_{1|1}) N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1
\end{aligned}
$$

We can solve the square of Gaussian distribution $N^2(y_2; HX_1, R)$ using equation (A.1) and obtain expression.

$$
E[(w_1^2)^2] = \frac{1}{\sqrt{4\pi R}} \int\int\int \frac{N(y_2; HX_1, R/2)}{N(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{1|1}, P_{1|1}) N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1
$$

$$
E[(w_1^2)^2] = \frac{1}{\sqrt{4\pi R}} \int\int\int \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{2|2}, P_{2|2}) N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1
$$

Now, using equation (A.2) yields

$$
\begin{aligned}
E[(w_1^2)^2] &= \frac{1}{\sqrt{4\pi R}} \sqrt{\frac{4\pi S_1^2}{R}} \int\int\int N(y_2; H\hat{x}_{1|1}, 2\frac{\tilde{S}_1 S_1}{R}) N(X_1; \hat{x}_{2|2}, P_{2|2}) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1
\end{aligned}
$$

Solving further and integrating over variable y_2, y_1 and X_1, we get

$$\begin{aligned}
E[(w_1^2)^2] &= \frac{S_1}{R}\int\int\int N(X_1;\hat{x}_{2|2},P_{2|2})N(y_2;H\hat{x}_{1|1},2\frac{\tilde{S_1}S_1}{R})N(y_1;H\hat{x}_{0|0},S_0)dX_1dy_2dy_1 \\
&= \frac{S_1}{R}\int\int N(X_1;\hat{x}_{2|2},P_{2|2})N(y_2;H\hat{x}_{1|1},2\frac{\tilde{S_1}S_1}{R})dX_1dy_2 \\
&= \frac{S_1}{R}\int N(X_1;\hat{x}_{2|2},P_{2|2})dX_1 \\
&= \frac{S_1}{R}
\end{aligned}$$

where integration of $\int N(y_1;H\hat{x}_{0|0},S_0)dy_1$, $\int N(y_2;H\hat{x}_{1|1},2\frac{\tilde{S_1}S_1}{R})dy_2$ and $\int N(X_1;\hat{x}_{2|2},P_{2|2})dX_1$ is equal to 1 in each case. Hence, we can write the simplified expression of the unconditional variance of w_1^2 as follows.

$$\begin{aligned}
Var(w_1^2) &= E[(w_1^2)^2] - (E[w_1^2])^2 \\
&= \frac{S_1}{R} - 1;
\end{aligned} \tag{A.4}$$

At $k = 3$, we have three samples with weights w_1^3, w_2^3 and w_3^3. The variance of w_3^3 is simply zero while the variance of w_1^3 and w_2^3 are derived below. The conditional variance of w_1^3 is given first.

$$\begin{aligned}
Var(w_1^3|Y^3) &= E[(w_1^3|Y^3)^2] - (E[w_1^3|Y^2])^2 \\
&= \int (w_1^3)^2 N(X_1;\hat{x}_{1|1},P_{1|1})dX_1 - 1 \\
&= \int \frac{N^2(y_3;HX_1,R)}{N^2(y_3;H\hat{x}_{2|2},S_2)}\frac{N^2(y_2;HX_1,R)}{N^2(y_2;H\hat{x}_{1|1},S_1)}N(X_1;\hat{x}_{1|1},P_{1|1})dX_1 - 1 \\
&= \frac{1}{4\pi R}\int \frac{N(y_3;HX_1,R/2)}{N^2(y_3;H\hat{x}_{2|2},S_2)}\frac{N(y_2;HX_1,R/2)}{N^2(y_2;H\hat{x}_{1|1},S_1)}N(X_1;\hat{x}_{1|1},P_{1|1})dX_1 - 1 \\
&= \frac{1}{4\pi R}\int \frac{N(y_3;HX_1,R/2)}{N^2(y_3;H\hat{x}_{2|2},S_2)}N(X_1;\bar{x}_{2|2},\bar{P}_{2|2})\frac{N(y_2;H\hat{x}_{1|1},S_1-R/2)}{N^2(y_2;H\hat{x}_{1|1},S_1)}dX_1 - 1 \\
&= \frac{1}{4\pi R}\int \frac{N(y_3;H\bar{x}_{2|2},H\bar{P}_{2|2}H'+R/2)}{N^2(y_3;H\hat{x}_{2|2},S_2)}N(X_1;\bar{x}_{3|3},\bar{P}_{3|3}) \\
&\times \frac{N(y_2;H\hat{x}_{1|1},S_1-R/2)}{N^2(y_2;H\hat{x}_{1|1},S_1)}dX_1 - 1
\end{aligned} \tag{A.5}$$

Now integrating over variable X_1, we obtain

$$Var(w_1^3|Y^3) = \frac{1}{4\pi R}\frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} - 1 \tag{A.6}$$

The unconditional variance of w_1^3 becomes

$$\begin{aligned}
Var(w_1^3) &= \frac{1}{4\pi R}\iiint \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\
&\times N(y_3; H\hat{x}_{2|2}, S_2)N(y_2; H\hat{x}_{1|1}, S_1)N(y_1; H\hat{x}_{0|0}, S_0)dy_3dy_2dy_1 - 1 \\
&= \frac{1}{4\pi R}\iiint \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N(y_3; H\hat{x}_{2|2}, S_2)}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N(y_2; H\hat{x}_{1|1}, S_1)} \\
&\times N(y_1; H\hat{x}_{0|0}, S_0)dy_3dy_2dy_1 - 1 \\
&= \frac{S_1}{R\sqrt{4\pi R}}\iiint \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N(y_3; H\hat{x}_{2|2}, S_2)}N(y_2; H\hat{x}_{1|1}, 2\tfrac{(S_1 - R/2)S_1}{R}) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0)dy_3dy_2dy_1 - 1
\end{aligned} \tag{A.7}$$

Now solving ratio of Gaussian densities in variable y_3 with different parameters, the above equation can be further simplified

$$\begin{aligned}
Var(w_1^3) &= \frac{S_1S_2}{R\sqrt{2R\Delta_2}}\iiint N(y_3; H\bar{x}_{2|2} - A_2\delta_2, S_2\Delta_2^{-1}\bar{S}_2)\exp(-\frac{\delta_2^2}{2}(A_2^2(\bar{S}_2^{-1} - S_2^{-1}) - S_2^{-1}) \\
&\times N(y_2; H\hat{x}_{1|1}, 2\tfrac{(S_1 - R/2)S_1}{R})N(y_1; H\hat{x}_{0|0}, S_0)dy_3dy_2dy_1 - 1 \\
&= \frac{S_1S_2}{R\sqrt{2R\Delta_2}}\iint \exp(-\frac{\delta_2^2}{2}((A_2^2(\bar{S}_2^{-1} - S_2^{-1}) - S_2^{-1})N(y_2; H\hat{x}_{1|1}, 2\tfrac{(S_1 - R/2)S_1}{R}) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0)dy_2dy_1 - 1 \\
&= \frac{S_1S_2^2\sqrt{2\pi(\bar{S}_2^{-1} - S_2^{-1})}}{R\sqrt{2|R\zeta_2H^2(k_1 - \bar{k}_1)^2(2 - S_2\bar{S}_2^{-1})|}}\iint N(y_2; H\hat{x}_{1|1}, 2\frac{(S_1 - R/2)S_1}{R}) \\
&\times N(y_2; H\hat{x}_{1|1}, \frac{(\bar{S}_2^{-1} - S_2^{-1})S_2^2}{H^2(k_1 - \bar{k}_1)^2(2 - S_2\bar{S}_2^{-1})})N(y_1; H\hat{x}_{0|0}, S_0)dy_2dy_1 - 1
\end{aligned} \tag{A.8}$$

Solving further, we obtain the simplified expression for $Var(w_1^3)$ as

$$\begin{aligned}
Var(w_1^3) &= \frac{S_1 S_2^2 \sqrt{2\pi(\bar{S}_2^{-1} - S_2^{-1})}}{R\sqrt{2|R\zeta_2 H^2 (k_1 - \bar{k}_1)^2 (2 - S_2\bar{S}_2^{-1})|}} \iint N(y_2; H\hat{x}_{1|1}, \mathrm{Y}_1) N(y_2; H\hat{x}_{1|1}, \mathrm{Y}_2) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0) dy_2 dy_1 - 1 \\
&= \frac{S_1 S_2^2 \sqrt{2\pi(\bar{S}_2^{-1} - S_2^{-1})}}{R\sqrt{2|R\zeta_2 H^2 (k_1 - \bar{k}_1)^2 (2 - S_2\bar{S}_2^{-1}) 2\pi(\mathrm{Y}_1 + \mathrm{Y}_2)|}} \iint N(y_2; H\hat{x}_{1|1}, \frac{\mathrm{Y}_1\mathrm{Y}_2}{\mathrm{Y}_1 + \mathrm{Y}_2}) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0) dy_2 dy_1 - 1 \\
&= \frac{S_1 S_2^2 \sqrt{2\pi(\bar{S}_2^{-1} - S_2^{-1})}}{R\sqrt{2|R\zeta_2 H^2 (k_1 - \bar{k}_1)^2 (2 - S_2\bar{S}_2^{-1}) 2\pi(\mathrm{Y}_1 + \mathrm{Y}_2)|}} - 1
\end{aligned} \tag{A.9}$$

where the parameters used are calculated below. These parameters result from solving the ratio of two Gaussian densities with different parameters.

$$\begin{aligned}
\zeta_2 &= S_2 - \bar{S}_2 \\
\delta_2 &= H\hat{x}_{2|2} - H\bar{x}_{2|2} \\
A_2 &= (\bar{S}_2^{-1} - S_2^{-1})^{-1} S_2^{-1} \\
\mathrm{Y}_1 &= 2\frac{(S_1 - R/2)S_1}{R} \\
\mathrm{Y}_2 &= \frac{(\bar{S}_2^{-1} - S_2^{-1}) S_2^2}{H^2 (k_1 - \bar{k}_1)^2 (2 - S_2\bar{S}_2^{-1})}
\end{aligned} \tag{A.10}$$

We find the the condition variance of $[w_2^3|Y^3]$ as below.

$$\begin{aligned}
Var(w_2^3|Y^3) &= E[(w_2^3|Y^3)^2] - (E[w_2^3|Y^3])^2 \\
&= \int (w_2^3)^2 N(X_2; \hat{x}_{2|2}, P_{2|2}) dX_2 - 1 \\
&= \int \frac{N^2(y_3; HX_2, R)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} N(X_2; \hat{x}_{2|2}, P_{2|2}) dX_2 - 1
\end{aligned} \tag{A.11}$$

Further solving, we obtain

$$\begin{aligned} Var(w_2^3|Y^3) &= \frac{1}{\sqrt{4\pi R}} \int \frac{N(y_3; HX_3, R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} N(X_2; \hat{x}_{2|2}, P_{2|2}) dX_2 - 1 \\ &= \frac{1}{\sqrt{4\pi R}} \int \frac{N(y_3; H\hat{x}_{2|2}, S_2 - R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} N(X_2; \bar{x}_{3|3}, \bar{P}_{3|3}) dX_2 - 1 \\ &= \frac{1}{\sqrt{4\pi R}} \frac{N(y_3; H\hat{x}_{2|2}, S_2 - R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} - 1 \end{aligned} \tag{A.12}$$

At $k = 4$, we have four samples with weights w_1^4, w_2^4 w_3^4 and w_4^4. The variance of w_4^4 is zero while variance of w_1^4, w_2^4 and w_3^4 are derived below. The conditional variance of w_1^4 is given first.

$$\begin{aligned} Var(w_1^4|Y^3) &= E[(w_1^4|Y^4)^2] - (E[w_1^4|Y^4])^2 \\ &= \int (w_1^4)^2 N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - 1 \\ &= \int \frac{N^2(y_4; HX_1, R)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N^2(y_3; HX_1, R)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \frac{N^2(y_2; HX_1, R)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\ &\times N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - 1 \end{aligned} \tag{A.13}$$

Solving the square of Guassians and further simplifying, we get

$$\begin{aligned} Var(w_1^4|Y^3) &= \frac{1}{(4\pi R)^{\frac{3}{2}}} \int \frac{N(y_4; HX_1, R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N(y_3; HX_1, R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \frac{N(y_2; HX_1, R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\ &\times N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - 1 \\ &= \frac{1}{(4\pi R)^{\frac{3}{2}}} \int \frac{N(y_4; HX_1, R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N(y_3; HX_1, R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} N(X_1; \bar{x}_{2|2}, \bar{P}_{2|2}) \\ &\times \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} dX_1 - 1 \\ &= \frac{1}{(4\pi R)^{\frac{3}{2}}} \int \frac{N(y_4; HX_1, R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} N(X_1; \bar{x}_{3|3}, \bar{P}_{3|3}) \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \\ &\times \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} dX_1 - 1 \end{aligned} \tag{A.14}$$

Further minimising the above expression, we obtain

$$
\begin{aligned}
&= \frac{1}{(4\pi R)^{\frac{3}{2}}} \int \frac{N(y_4; H\bar{x}_{3|3}, H\bar{P}_{3|3}H' + R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} N(X_1; \bar{x}_{4|4}, \bar{P}_{4|4}) \\
&\times \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} dX_1 - 1 \\
&= \frac{1}{(4\pi R)^{\frac{3}{2}}} \frac{N(y_4; H\bar{x}_{3|3}, H\bar{P}_{3|3}H' + R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \\
&\times \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} - 1 \\
&= \frac{1}{(4\pi R)^{\frac{3}{2}}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \prod_{i=3}^{4} \frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} - 1
\end{aligned}
\tag{A.15}
$$

The conditional variance of w_2^4 for a sample drawn at $k = 2$ and updated at $k = 4$ is derived in below set of equations. Using the formula of variance of a random variable in terms of expected value, we write

$$
\begin{aligned}
Var(w_2^4|Y^4) &= E[(w_2^4|Y^4)^2] - (E[w_2^4|Y^4])^2 \\
&= \int (w_2^4)^2 N(X_2; \hat{x}_{2|2}, P_{2|2}) dX_2 - 1
\end{aligned}
$$

Substituting the expression of w_2^4, the above expression becomes

$$
Var(w_2^4|Y^4) = \int \frac{N^2(y_4; HX_2, R)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N^2(y_3; HX_3, R)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} N(X_2; \hat{x}_{2|2}, P_{2|2}) dX_2 - 1
$$

Solving the product of Gaussian, $Var(w_2^4|Y^4)$ can be expressed as

$$
\begin{aligned}
Var(w_2^4|Y^4) &= \frac{1}{4\pi R} \int \frac{N(y_4; HX_2, R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N(y_3; HX_2, R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} N(X_2; \hat{x}_{2|2}, P_{2|2}) dX_2 - 1 \\
&= \frac{1}{4\pi R} \int \frac{N(y_4; HX_2, R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} N(X_2; \bar{x}_{3|3}, \bar{P}_{3|3}) \frac{N(y_3; H\hat{x}_{2|2}, S_2 - R/2)}{N^2(y_2; H\hat{x}_{2|2}, S_2)} dX2 - 1 \\
&= \frac{1}{4\pi R} \int \frac{N(y_4; H\bar{x}_{3|3}, H\bar{P}_{3|3}H' + R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} N(X_2; \bar{x}_{4|4}, \bar{P}_{4|4}) \\
&\times \frac{N(y_3; H\hat{x}_{2|2}, S_2 - R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} dX_2 - 1
\end{aligned}
$$

Integrating over variable X_2, we get the below simplified form of $Var(w_2^4|Y^4)$

$$= \frac{1}{4\pi R} \frac{N(y_4; H\bar{x}_{3|3}, H\bar{P}_{3|3}H' + R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N(y_3; H\hat{x}_{2|2}, S_2 - R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} - 1 \tag{A.16}$$

Now, we find the the condition variance of $[w_3^4|Y^4]$ as

$$\begin{aligned} Var(w_3^4|Y^4) &= E[(w_3^4|Y^4)^2] - (E[w_3^4|Y^4])^2 \\ &= \int (w_3^4)^2 N(X_3; \hat{x}_{3|3}, P_{3|3}) dX_3 - 1 \\ &= \int \frac{N^2(y_4; HX_3, R)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} N(X_3; \hat{x}_{3|3}, P_{3|3}) dX_3 - 1 \\ &= \frac{1}{\sqrt{4\pi R}} \int \frac{N(y_4; HX_3, R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} N(X_3; \hat{x}_{3|3}, P_{3|3}) dX_3 - 1 \\ &= \frac{1}{\sqrt{4\pi R}} \int \frac{N(y_4; H\hat{x}_{3|3}, S_3 - R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} N(X_2; \bar{x}_{3|3}, \bar{P}_{3|3}) dX_2 \\ &= \frac{1}{\sqrt{4\pi R}} \frac{N(y_4; H\hat{x}_{3|3}, S_3 - R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \end{aligned} \tag{A.17}$$

At time k, we have k samples with weights w_1^k, w_2^k ,w_3^k, upto w_k^k. Using induction and the above results, we can write the conditional variance of the above weights as

$$\begin{aligned} \Upsilon_1^k &= \frac{1}{(4\pi R)^{\frac{k-1}{2}}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \prod_{i=3}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_{i-1})} - 1 \\ \Upsilon_2^k &= \frac{1}{(4\pi R)^{\frac{k-2}{2}}} \frac{N(y_3; H\hat{x}_{2|2}, S_2 - R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \prod_{i=4}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_{i-1})} - 1 \\ \Psi_3^k &= \frac{1}{(4\pi R)^{\frac{k-3}{2}}} \frac{N(y_4; H\hat{x}_{3|3}, S_3 - R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \prod_{i=5}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_{i-1})} - 1 \end{aligned}$$

.

.

$$\Psi_{k-1}^k = \frac{1}{(4\pi R)^{\frac{1}{2}}} \frac{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1} - R/2)}{N^2(y_k; H\hat{x}_{k-1|k-1}, S_{k-1})} - 1$$

Using the above expressions, the general form becomes.

$$\begin{aligned}\Psi_n^k = &\frac{1}{\sqrt[k-n]{4\pi R}} \frac{N(y_{n+1}; H\hat{x}_{n|n}, S_n - R/2)}{N^2(y_{n+1}; H\hat{x}_{n|n}, S_n)} \\ &\times \prod_{i=n+2}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} - 1\end{aligned} \tag{A.18}$$

Where Ψ_n^k is conditional variance of weight $w_n^k|Y^k$, $S_{k-1} = HP_{k|k-1}H' + R$, $\bar{S}_{i-1} = H\bar{P}_{i|i-1}H' + R/2$ and $n \leqslant k$ represents the time at which samples were drawn and k is current time index.

A.3 Conditional Expected Value of the Estimate (Section 5.4.2)

In this appendix, we derive closed form expressions for the expected values of our proposed estimates conditional on measurements. We use induction for its derivations. At time $k = 2$,

$$\begin{aligned}E[\Delta_2] &= E\left[\frac{w_1^2 X_1 + w_2^2 X_2}{2}\right] \\ &= \frac{1}{2} E\left[w_1^2 X_1 + w_2^2 X_2\right] \\ &= \frac{1}{2}\left[E[w_1^2 X_1] + E[w_2^2 X_2]\right] \\ &= \frac{1}{2}\left[E[l_2(X_1)\frac{C_1^1}{C_1^2} X_1] + E[X_2\right] \\ &= \frac{1}{2}\left[\int l_2(X_1)\frac{C_1^1}{C_1^2} X_1 \frac{l_1(X_1)\pi_0(X_1)}{C_1^1} dX_1 + \int X_2 \pi_2(X_2) dX_2\right] \\ &= \frac{1}{2}\left[\int X_1 \frac{l_2(X_1) l_1(X_1)\pi_0(X_1)}{C_1^2} dX_1 + \int X_2 \pi_2(X_2) dX_2\right] \\ &= \frac{1}{2}\left[\int X_1 \pi_2(X_1) dX_1 + \int X_2 \pi_2(X_2) dX_2\right] \\ &= \frac{1}{2}\left[\hat{x}_{2|2} + \hat{x}_{2|2}\right] \\ &= \hat{x}_{2|2}\end{aligned} \tag{A.19}$$

At time $k = 3$,

$$
\begin{aligned}
E[\Delta_3] &= E\left[\frac{w_1^3 X_1 + w_2^3 X_2 + w_3^3 X_3}{3}\right] \\
&= \frac{1}{3}\Big\{E[w_1^3 X_1] + E[w_2^3 X_2] + E[w_3^3 X_3]\Big\} \\
&= \frac{1}{3}\Big\{\int l_3(X_1) w_1^2 \frac{C_1^2}{C_1^3} X_1 \frac{l_1(X_1)\pi_0(X_1)}{C_1^1} dX_1 + \int l_3(X_2) w_2^2 \frac{C_2^2}{C_2^3} X_2 \frac{l_2(X_2) l_1(X_2)\pi_0(X_2)}{C_2^2}) dX_2 \\
&+ \int X_3 \pi_3(X_3) dX_3\Big\} \\
&= \frac{1}{3}\Big\{\int l_3(X_1) l_2(X_1) \frac{C_1^1}{C_1^2}\frac{C_1^2}{C_1^3} X_1 \frac{l_1(X_1)\pi_0(X_1)}{C_1^1} dX_1 \\
&+ \int l_3(X_2) w_2^2 \frac{C_2^2}{C_2^3} X_2 \frac{l_2(X_2) l_1(X_2)\pi_0(X_2)}{C_2^2}) dX_2 + \int X_3 \pi_3(X_3) dX_3\Big\} \\
&= \frac{1}{3}\Big\{\int X_1 \frac{l_3(X_1) l_2(X_1) l_1(X_1)\pi_0(X_1)}{C_1^3} dX_1 + \int X_2 \frac{l_3(X_2) l_2(X_2) l_1(X_2)\pi_0(X_2)}{C_2^3} dX_2 \\
&+ \int X_3 \pi_3(X_3) dX_3\Big\} \\
&= \frac{1}{3}\Big\{\int X_1 \pi_3(X_1) dX_1 + \int X_2 \pi_2(X_2) dX_2 + \int X_3 \pi_3(X_3) dX_3\Big\} \\
&= \hat{x}_{3|3}
\end{aligned}
\tag{A.20}
$$

At time $k = 4$

$$
\begin{aligned}
E[\Delta_4] &= E\left[\frac{w_1^4 X_1 + w_2^4 X_2 + w_3^4 X_3 + w_4^4 X_4}{4}\right] \\
&= \frac{1}{4}\Big\{E[l_4(X_1) w_1^3 \frac{C_1^3}{C_1^4} X_1] + E[l_4(X_2) w_2^3 \frac{C_2^3}{C_2^4} X_2] + E[l_4(X_3) w_3^4 \frac{C_3^3}{C_3^4} X_3] + E[X_4]\Big\} \\
&= \frac{1}{4}\Big\{\int X_1 \frac{l_4(X_1) l_3(X_1) l_2(X_1) l_1(X_1)\pi_0(X_1)}{C_1^4} dX_1 + \int X_2 \frac{l_4(X_2) l_3(X_2) l_2(X_2) l_1(X_2)\pi_0(X_2)}{C_2^4} \\
&dX_2 + \int X_3 \frac{l_4(X_3) l_3(X_3) l_2(X_3) l_1(X_3)\pi_0(X_3)}{C_3^4} dX_3 + \int X_4 \pi_4(X_4) dX_4\Big\} \\
&= \frac{1}{4}\Big\{\int X_1 \pi_4(X_1) dX_1 + \int X_2 \pi_4(X_2) dX_2] + \int X_3 \pi_4(X_3) dX_3 + \int X_4 \pi_4(X_4) dX_4\Big\} \\
&= \hat{x}_{4|4}
\end{aligned}
\tag{A.21}
$$

At time k, using induction, we derive results for $E[\Delta_k]$

$$\begin{aligned}
E[\Delta_k] &= E\left[\frac{w_1^k X_1 + w_2^k X_2 + w_3^k X_3 + ... + w_k^k X_k}{k}\right] \\
&= \frac{1}{k}\Big\{E\left[w_1^k X_1 + w_2^k X_2 + w_3^k X_3 + ... + w_k^k X_k\right]\Big\} \\
&= \frac{1}{k}\Big\{E[w_1^k X_1] + E[w_2^k X_2] + E[w_3^k X_3] + ... + E[w_k^k X_k]\Big\} \\
&= \frac{1}{k}\Big\{\int X_1 \pi_k(X_1) dX_1 + \int X_2 \pi_k(X_2) dX_2 + \int X_3 \pi_k(X_3) dX_3 + ... + \int X_k \pi_k(X_k) dX_k\Big\}
\end{aligned} \tag{A.22}$$

Evaluating the individual expectation of the terms in above equation, we can write the expression for $E[\Delta_k]$ as

$$\begin{aligned}
E[\Delta_k] &= \frac{1}{k}\left[\hat{x}_{k|k} + \hat{x}_{k|k} + ... + \hat{x}_{k|k}\right] \\
&= \hat{x}_{k|k}
\end{aligned} \tag{A.23}$$

A.4 Unconditional Expected Value of the Estimator

We include the derivations of unconditional expected values of the proposed estimator. Again, we use induction to derive the general formulae. At time $k = 2$, we solve for the expectation over $x, y1, y2$ for $k = 2$ and obtain

$$\begin{aligned}
E[\Delta_2] &= E\left[\frac{w_1^2 X_1 + w_2^2 X_2}{2}\right] \\
&= \frac{1}{2} E[w_1^2 X_1 + w_2^2 X_2] \\
&= \frac{1}{2}\{E[w_1^2 X_1] + E[w_2^2 X_2]\}
\end{aligned}$$

We derive the closed form solution of $E[w_1^2 X_1]$ and $E[w_2^2 X_2]$ separately and then we combine them for the final result.

$$E[w_1^2 X_1] = \int\int\int w_1^2 X_1 p(X_1, y2, y1) dX_1 dy_2 dy_1$$

Considering everything linear and Gaussian, the joint distribution $p(X_1, y2, y1)$ is given as below using a Kalman filter framework.

$$
\begin{aligned}
p(X_1, y2, y1) &= N(X_1; \hat{x}_{1|1}, P_{1|1}) N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) \\
S_1 &= HP_{1|1}H' + R \\
C_1 &= N(y_1; H\hat{x}_{0|0}, S_0) \\
C_2 &= N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; \hat{x}_{0|0}, S_0) \\
w_1^2 &= \frac{N(y_2; Hx_1, R)}{N(y_2; H\hat{x}_{1|1}, S_1)}
\end{aligned}
$$

Putting these terms back we get

$$
\begin{aligned}
E[w_1^2 X_1] &= \int\int\int X_1 \frac{N(y_2; HX_1, R)}{N(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{1|1}, P_{1|1}) N(y_2; H\hat{x}_{1|1}, S_1) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1 \\
&= \int\int\int X_1 \frac{N(y_2; H\hat{x}_{1|1}, S_1)}{N(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{2|2}, P_{2|2}) N(y_2; H\hat{x}_{1|1}, S_1) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1 \\
&= \int\int\int X_1 N(X_1; \hat{x}_{2|2}, P_{2|2}) N(y_2; H\hat{x}_{1|1}, S_1) \\
&\times N(y_1; H\hat{x}_{0|0}, S_0) dX_1 dy_2 dy_1 \\
&= \int\int \hat{x}_{2|2} N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dy_2 dy_1 \\
&= \int\int (\hat{x}_{1|1} + K_1 y_2 - K_1 H\hat{x}_{1|1}) N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dy_2 dy_1
\end{aligned}
$$

Putting $\hat{x}_{2|2} = \hat{x}_{1|1} + K_1 y_2 - K_1 H\hat{x}_{1|1}$ and integrating first over y_2.

$$
\begin{aligned}
E[w_1^2 X_1] &= \int\int (\hat{x}_{1|1} + K_1 y_2 - K_1 H\hat{x}_{1|1}) N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dy_2 dy_1 \\
&= \int\int \hat{x}_{1|1} N(y_1; H\hat{x}_{0|0}, S_0) dy_1
\end{aligned}
$$

Putting $\hat{x}_{1|1} = \hat{x}_{0|0} + K_0 y_1 - K_0 H \hat{x}_{0|0}$ and integrating first over variable y_2.

$$\begin{aligned} &= \int (\hat{x}_{0|0} + K_0 y_1 - K_0 H \hat{x}_{0|0}) N(y_1; H\hat{x}_{0|0}, S_0) dy_1 \\ &= \hat{x}_{0|0} \end{aligned}$$

Similarly, $E[w_2^2 X_2]$ when $w_2^2 = 1$ can be derived as below.

$$\begin{aligned} E[w_2^2 X_2] &= \int\int\int X_2 N(X_2; \hat{x}_{2|2}, P_{2|2}) N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dX_2 dy_2 dy_1 \\ &= \int\int\int X_2 N(X_2; \hat{x}_{2|2}, P_{2|2}) N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dX_2 dy_2 dy_1 \\ &= \int\int \hat{x}_{2|2} N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dy_2 dy_1 \\ &= \int\int (\hat{x}_{1|1} + K_1 y_2 - K_1 H \hat{x}_{1|1}) N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dy_2 dy_1 \end{aligned}$$

Putting $\hat{x}_{2|2} = \hat{x}_{1|1} + K_1 y_2 - K_1 H \hat{x}_{1|1}$ and integrating first over y_2.

$$\begin{aligned} E[w_2^2 X_2] &= \int\int (\hat{x}_{1|1} + K_1 y_2 - K_1 H \hat{x}_{1|1}) N(y_2; H\hat{x}_{1|1}, S_1) N(y_1; H\hat{x}_{0|0}, S_0) dy_2 dy_1 \\ &= \int\int \hat{x}_{1|1} N(y_1; H\hat{x}_{0|0}, S_0) dy_1 \end{aligned}$$

Putting $\hat{x}_{1|1} = \hat{x}_{0|0} + K_0 y_1 - K_0 H \hat{x}_{0|0}$ and integrating first over y_1.

$$\begin{aligned} &= \int (\hat{x}_{0|0} + K_0 y_1 - K_0 H \hat{x}_{0|0}) N(y_1; H\hat{x}_{0|0}, S_0) dy_1 \\ &= \hat{x}_{0|0} \end{aligned}$$

Substituting the above results back, yields

$$\begin{aligned} E[\hat{x}_2] &= E\left[\frac{w_1^2 X_1 + w_2^2 X_2}{2}\right] \\ &= \frac{1}{2} E\left[w_1^2 X_1 + w_2^2 X_2\right] \\ &= \frac{1}{2}\Big\{E[w_1^2 X_1] + E[w_2^2 X_2]\Big\} \\ &= \frac{1}{2}\Big\{\hat{x}_{0|0} + \hat{x}_{0|0}\Big\} \\ &= \hat{x}_{0|0} \end{aligned}$$

A.5 Conditional Variance of the proposed PBDF (Section 5.4.2)

In this appendix, we derive the variance of the proposed estimator expressed as $Var[\Delta_k] = \frac{1}{k^2}\sum_{n=1}^{k}\Theta_n^k$. We first derive the closed form expressions for the conditional variance of the product of sample and weights $w_n^k X_n$ expressed as Θ_n^k. The general form of the Θ_n^k in terms of expected value can be written as

$$\Theta_n^k = E[(w_n^k X_n|Y^k)^2] - (E[w_n^k X_n|Y^k])^2 \tag{A.24}$$

The expected value of $w_1^k X_1$ conditional on all measurements received upto time k is $\hat{x}_{k|k}$ and the conditional variance of $w_1^k X_1$ is being derived as below by using the induction procedure. We start the exploring the expressions of $\Theta_1^2, \Theta_1^3, \Theta_1^4$ and the derive the general expression Θ_1^k The variance of the $w_1^k X_1$ when $k = 2$

$$\begin{aligned}
\Theta_1^2 &= E[(w_1^2 X_1|Y^2)^2] - (E[w_1^2 X_1|Y^2])^2 \\
&= \int (w_1^2)^2 X_1^2 N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - \hat{x}_{2|2}^2 \\
&= \int X_1^2 \frac{N^2(y_2; HX_1, R)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - \hat{x}_{2|2}^2 \\
&= \frac{1}{\sqrt{4\pi R}} \int X_1^2 \frac{N(y_2; HX_1, R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - \hat{x}_{2|2}^2 \\
&= \frac{1}{\sqrt{4\pi R}} \int \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} X_1^2 N(X_1; \bar{x}_{2|2}, \bar{P}_{2|2}) dX_1 - \hat{x}_{2|2}^2 \\
&= \frac{1}{\sqrt{4\pi R}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \int X_1^2 N(X_1; \bar{x}_{2|2}, \bar{P}_{2|2}) dX_1 - \hat{x}_{2|2}^2 \\
&= \frac{1}{\sqrt{4\pi R}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \{\bar{P}_{2|2} + \bar{x}_{2|2}^2\} - \hat{x}_{2|2}^2
\end{aligned} \tag{A.25}$$

The variance of the $w_1^k X_1$ when $k = 3$

$$\begin{aligned}
\Theta_1^3 &= E[(w_1^3 X_1|Y^3)^2] - (E[w_1^3 X_1|Y^2])^2 \\
&= \int (w_1^3)^2 X_1^2 N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - \hat{x}_{3|3}^2 \\
&= \int \frac{N^2(y_3; HX_1, R)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \frac{N^2(y_2; HX_1, R)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} X_1^2 N(X_1; \hat{x}_{1|1}, P_{1|1}) dX_1 - \hat{x}_{3|3}^2
\end{aligned} \tag{A.26}$$

Rearranging and solving the products of the Gaussian, we further solve as below.

$$\begin{aligned}
\Theta_1^3 &= \frac{1}{4\pi R}\int \frac{N(y_3; HX_1, R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)}\frac{N(y_2; HX_1, R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}X_1^2 N(X_1; \hat{x}_{1|1}, P_{1|1})dX_1 - \hat{x}_{3|3}^2 \\
&= \frac{1}{4\pi R}\int \frac{N(y_3; HX_1, R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)}X_1^2 N(X_1; \bar{x}_{2|2}, \bar{P}_{2|2})\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}dX_1 - \hat{x}_{3|3}^2 \\
&= \frac{1}{4\pi R}\int \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)}X_1^2 N(X_1; \bar{x}_{3|3}, \bar{P}_{3|3}) \\
&\times \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}dX_1 - \hat{x}_{3|3}^2 \\
&= \frac{1}{4\pi R}\frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\
&\times \int X_1^2 N(X_1; \bar{x}_{3|3}, \bar{P}_{3|3})dX_1 - \hat{x}_{3|3}^2 \\
&= \frac{1}{4\pi R}\frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}\{\bar{P}_{3|3} + \bar{x}_{3|3}^2\} - \hat{x}_{3|3}^2
\end{aligned} \tag{A.27}$$

The variance of the $w_1^k X_1$ when $k = 4$

$$\begin{aligned}
\Theta_1^4 &= E[(w_1^4 X_1|Y^4)^2] - (E[w_1^4 X_1|Y^4])^2 \\
&= \int (w_1^4)^2 X_1^2 N(X_1; \hat{x}_{1|1}, P_{1|1})dX_1 - \hat{x}_{4|4}^2
\end{aligned} \tag{A.28}$$

Replacing the the expression for w_1^4, we solve further as

$$\begin{aligned}
&= \int \frac{N^2(y_4; HX_1, R)}{N^2(y_4; H\hat{x}_{3|3}, S_3)}\frac{N^2(y_3; HX_1, R)}{N^2(y_3; H\hat{x}_{2|2}, S_2)}\frac{N^2(y_2; HX_1, R)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\
&\times X_1^2 N(X_1; \hat{x}_{1|1}, P_{1|1})dX_1 - \hat{x}_{4|4}^2 \\
&= \frac{1}{(4\pi R)^{\frac{3}{2}}}\int \frac{N^2(y_4; HX_1, R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)}\frac{N^2(y_3; HX_1, R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \\
&\times \frac{N^2(y_2; HX_1, R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}X_1^2 N(X_1; \hat{x}_{1|1}, P_{1|1})dX_1 - \hat{x}_{4|4}^2
\end{aligned} \tag{A.29}$$

Using the Gaussian product theorem and solving the above expressions further, we go as below.

$$
\begin{aligned}
\Theta_1^4 &= \frac{1}{(4\pi R)^{\frac{3}{2}}} \int \frac{N(y_4; HX_1, R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N(y_3; HX_1, R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} X_1^2 N(X_1; \bar{x}_{2|2}, \bar{P}_{2|2}) \\
&\times \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} dX_1 - \hat{x}_{4|4}^2 \\
&= \frac{1}{(4\pi R)^{\frac{3}{2}}} \int \frac{N(y_4; H\bar{x}_{3|3}, H\bar{P}_{3|3}H' + R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} X_1^2 N(X_1; \bar{x}_{4|4}, \bar{P}_{4|4}) \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \\
&\times \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} dX_1 - \hat{x}_{4|4}^2 \\
&= \frac{1}{(4\pi R)^{\frac{3}{2}}} \frac{N(y_4; H\bar{x}_{3|3}, H\bar{P}_{3|3}H' + R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \frac{N(y_3; H\bar{x}_{2|2}, H\bar{P}_{2|2}H' + R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \\
&\times \int X_1^2 N(X_1; \bar{x}_{4|4}, \bar{P}_{4|4}) dX_1 - \hat{x}_{4|4}^2 \\
&= \frac{1}{(4\pi R)^{\frac{3}{2}}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \prod_{i=3}^{4} \frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \{\bar{P}_{4|4} + \bar{x}_{4|4}^2\} - \hat{x}_{4|4}^2
\end{aligned}
\tag{A.30}
$$

Similarly, the variance of the $w_1^k X_1$ at time $k-1$ is written by using induction procedure as below.

$$
\begin{aligned}
\Theta_1^{k-2} &= \frac{1}{(4\pi R)^{\frac{k-3}{2}}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \prod_{i=3}^{k-2} \frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\
&\quad \times \left\{ \bar{P}_{k-2|k-2} + \bar{x}_{k-2|k-2}^2 \right\} - \hat{x}_{k-2|k-2}^2 \\
\Theta_1^{k-1} &= \frac{1}{(4\pi R)^{\frac{k-2}{2}}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \prod_{i=3}^{k-1} \frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\
&\quad \times \left\{ \bar{P}_{k-1|k-1} + \bar{x}_{k-1|k-1}^2 \right\} - \hat{x}_{k-1|k-1}^2 \\
\Theta_1^{k} &= \frac{1}{(4\pi R)^{\frac{k-1}{2}}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \prod_{i=3}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\
&\quad \times \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2
\end{aligned}
\tag{A.31}
$$

Now, We generalise the results for samples drawn at time n and their respective weights are updated at time k where $n < k$.

$$\Theta_1^k = \frac{1}{(4\pi R)^{\frac{k-1}{2}}} \frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)} \prod_{i=3}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2$$

$$\Theta_2^k = \frac{1}{(4\pi R)^{\frac{k-2}{2}}} \frac{N(y_3; H\hat{x}_{2|2}, S_2 - R/2)}{N^2(y_3; H\hat{x}_{2|2}, S_2)} \prod_{i=4}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2$$

$$\Theta_3^k = \frac{1}{(4\pi R)^{\frac{k-3}{2}}} \frac{N(y_4; H\hat{x}_{3|3}, S_3 - R/2)}{N^2(y_4; H\hat{x}_{3|3}, S_3)} \prod_{i=5}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2$$

.

.

.

$$\Theta_{k-1}^k = \frac{1}{(4\pi R)^{\frac{1}{2}}} \frac{N(y_k; H\hat{x}_{k-1|k-1}, S_{k-1} - R/2)}{N^2(y_k; H\hat{x}_{k-1|k-1}, S_{k-1})} \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2 \tag{A.32}$$

Using above expressions, the general form is as below.

$$\begin{aligned} \Theta_n^k &= E[(w_n^k X_n | Y^k)^2] - (E[w_n^k X_n | Y^k])^2 \\ \Theta_n^k &= \frac{1}{(4\pi R)^{\frac{k-n}{2}}} \frac{N(y_{n+1}; H\hat{x}_{n|n}, S_n - R/2)}{N^2(y_{n+1}; H\hat{x}_{n|n}, S_n)} \prod_{i=n+2}^{k} \frac{N(y_i; H\bar{x}_{i-1|i-1}, \bar{S}_{i-1})}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\ &\quad \times \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2 \end{aligned} \tag{A.33}$$

The above equation is complex and to understand we write the recursive form of the above equation in the next section.

A.6 Recursive form of Θ_n^k

In this Appendix, we write the recursion form of Θ_n^k. Firstly, we derive the recursive expression for Θ_1^k and then generalise the expression to the case of Θ_n^k.Now using the above expressions, we can derive $Var(w_1^k | Y^k)$ as below.

$$\begin{aligned}\Theta_1^k &= \frac{1}{(4\pi R)^{\frac{k-1}{2}}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}\prod_{i=3}^{k}\frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\ &\times\left\{\bar{P}_{k|k} + \bar{x}_{k|k}^2\right\} - \hat{x}_{k|k}^2 \\ &= \frac{1}{(4\pi R)^{k-2/2}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}\prod_{i=3}^{k-1}\frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\ &\frac{1}{(4\pi R)^{\frac{1}{2}}}\frac{N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{k-1|k-1}H' + R/2))}{N^2(y_k; H\hat{x}_{k-1|k-1}, S_{k-1})}\left\{\bar{P}_{k|k} + \bar{x}_{k|k}^2\right\} - \hat{x}_{k|k}^2\end{aligned} \tag{A.34}$$

Re-arranging the expression for Θ_1^{k-1}, we get below form to replace in above equation.

$$\begin{aligned}\Theta_1^{k-1} &= \frac{1}{(4\pi R)^{\frac{k-2}{2}}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}\prod_{i=3}^{k-1}\frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\ &\times\left\{\bar{P}_{k|k} + \bar{x}_{k|k}^2\right\} - \hat{x}_{k-1|k-1}^2 \\ \Theta_1^{k-1} + \hat{x}_{k-1|k-1}^2 &= \frac{1}{(4\pi R)^{\frac{k-2}{2}}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}\prod_{i=3}^{k-1}\frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)} \\ &\times\left\{\bar{P}_{k|k} + \bar{x}_{k|k}^2\right\} \\ \frac{\Theta_1^{k-1} + \hat{x}_{k-1|k-1}^2}{\bar{P}_{k-1|k-1} + \bar{x}_{k-1|k-1}^2} &= \frac{1}{(4\pi R)^{\frac{k-2}{2}}}\frac{N(y_2; H\hat{x}_{1|1}, S_1 - R/2)}{N^2(y_2; H\hat{x}_{1|1}, S_1)}\prod_{i=3}^{k-1}\frac{N(y_i; H\bar{x}_{i-1|i-1}, H\bar{P}_{i-1|i-1}H' + R/2)}{N^2(y_i; H\hat{x}_{i-1|i-1}, S_i)}\end{aligned} \tag{A.35}$$

Using Equations (A.35), we get below recursion form for Θ_1^k

$$\begin{aligned}\Theta_1^k &= \frac{1}{(4\pi R)^{\frac{1}{2}}}\frac{N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{k-1|k-1}H' + R/2))}{N^2(y_k; H\hat{x}_{k-1|k-1}, S_{k-1})}\left\{\frac{\Theta_1^{k-1} + \hat{x}_{k-1|k-1}^2}{\bar{P}_{k-1|k-1} + \bar{x}_{k-1|k-1}^2}\right\} \\ &\times\left\{\bar{P}_{k|k} + \bar{x}_{k|k}^2\right\} - \hat{x}_{k|k}^2 \\ &= \sqrt{\frac{S_{k-1}}{R}}\frac{N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{k-1|k-1}H' + R/2))}{N(y_k; H\hat{x}_{k-1|k-1}, \frac{S_{k-1}}{2})}\left\{\frac{\Theta_1^{k-1} + \hat{x}_{k-1|k-1}^2}{\bar{P}_{k-1|k-1} + \bar{x}_{k-1|k-1}^2}\right\} \\ &\times\left\{\bar{P}_{k|k} + \bar{x}_{k|k}^2\right\} - \hat{x}_{k|k}^2\end{aligned} \tag{A.36}$$

Using the above procedure, we can write recursive forms for Θ_2^k, Θ_3^k and generalise it to the expression of Θ_n^k as below.

$$\begin{aligned}\Theta_n^k = \sqrt{\frac{S_{k-1}}{R}} \frac{N(y_k; H\bar{x}_{k-1|k-1}, H\bar{P}_{i-1|i-1}H' + R/2))}{N(y_k; H\hat{x}_{k-1|k-1}, \frac{S_{k-1}}{2})} \left\{ \frac{\Theta_n^{k-1} + \hat{x}_{k-1|k-1}^2}{\bar{P}_{k-1|k-1} + \bar{x}_{k-1|k-1}^2} \right\} \\ \times \left\{ \bar{P}_{k|k} + \bar{x}_{k|k}^2 \right\} - \hat{x}_{k|k}^2\end{aligned} \tag{A.37}$$

Appendix B

Convergence of the Proposed Estimator

B.1 Variance of Estimator in Terms of Weight Variance

In this part of Appendix, we derive an expression for Θ_n^k as a function of Ψ_n^k which implies the dependence of the estimator variance on the variance of the weights.

$$\begin{aligned}\Theta_n^k = &\frac{1}{(4\pi R)^{\frac{k-n}{2}}}\frac{N(y_{n+1};H\hat{x}_{n|n},S_n-R/2)}{N^2(y_{n+1};H\hat{x}_{n|n},S_n)}\prod_{i=n+2}^{k}\frac{N(y_i;H\bar{x}_{i-1|i-1},\bar{S}_{i-1})}{N^2(y_i;H\hat{x}_{i-1|i-1},S_i)}\\ &\times\{\bar{P}_{k|k}+\bar{x}_{k|k}^2\}-\hat{x}_{k|k}^2\end{aligned} \tag{B.1}$$

The re-arrangement of variance of sample weights is done as below.

$$\begin{aligned}\Psi_n^k &= \frac{1}{(4\pi R)^{\frac{k-n}{2}}}\frac{N(y_{n+1};H\hat{x}_{n|n},S_n-R/2)}{N^2(y_{n+1};H\hat{x}_{n|n},S_n)}\prod_{i=n+2}^{k}\frac{N(y_i;H\bar{x}_{i-1|i-1},\bar{S}_{i-1})}{N^2(y_i;H\hat{x}_{i-1|i-1},S_i)}-1\\ \Psi_n^k+1 &= \frac{1}{(4\pi R)^{\frac{k-n}{2}}}\frac{N(y_{n+1};H\hat{x}_{n|n},S_n-R/2)}{N^2(y_{n+1};H\hat{x}_{n|n},S_n)}\prod_{i=n+2}^{k}\frac{N(y_i;H\bar{x}_{i-1|i-1},\bar{S}_{i-1})}{N^2(y_i;H\hat{x}_{i-1|i-1},S_i)}\end{aligned} \tag{B.2}$$

By replacing above expression in equation (B.1), we get the form of Θ_n^k as a function of Ψ_n^k.

$$\begin{aligned}\Theta_n^k &= \frac{1}{(4\pi R)^{\frac{k-n}{2}}}\frac{N(y_{n+1};H\hat{x}_{n|n},S_n-R/2)}{N^2(y_{n+1};H\hat{x}_{n|n},S_n)}\prod_{i=n+2}^{k}\frac{N(y_i;H\bar{x}_{i-1|i-1},\bar{S}_{i-1})}{N^2(y_i;H\hat{x}_{i-1|i-1},S_i)}\{\bar{P}_{k|k}+\bar{x}_{k|k}^2\}-\hat{x}_{k|k}^2\\ &= \{\Psi_n^k+1\}\{\bar{P}_{k|k}+\bar{x}_{k|k}^2\}-\hat{x}_{k|k}^2\end{aligned}$$

B.2 Kalman Filter Expressions

In this appendix, we derive expressions for the terms S_k, $\bar{S}_k$, P_k and $\bar{P}_k$ as functions of time index k and initial values of x_0 and P_0. We assume the simple case of linear Gaussian parameter estimation with $H = 1$, and only deal with one dimension, i.e., scalar parameter and measurements.

$$
\begin{aligned}
P1 &= P_0 - K_1 P_0 = P_0 - \frac{P_0^2}{P_0 + R} = \frac{P_0^2 + P_0 R - P_0^2}{P_0 + R} \\
P_1 &= \frac{P_0 R}{P_0 + R} \\
P_2 &= P_1 - K_2 P_1 = P_1 - \frac{P_1^2}{P_1 + R} = \frac{P_1^2 + P_1 R - P_1^2}{P_1 + R} \\
P_2 &= \frac{P_1 R}{P_1 + R}
\end{aligned}
$$

Now putting expression of P_1, we get the following form of P_2

$$
P_2 = \frac{\frac{P_0 R^2}{P_0 + R}}{\frac{P_0 R}{P_0 + R} + R} = \frac{P_o R}{2P_0 + R}
$$

Similarly, we can write general expression using induction procedure as above.

$$
P_k = \frac{P_o R}{kP_0 + R} \tag{B.3}
$$

We derive the S_k, $\bar{P}_k$ and $\bar{S}_k$ using above induction procedure and results are written

in below set of equations.

$$\begin{aligned} S_k &= \frac{kP_oR + R^2}{(k-1)P_0 + R} \\ \bar{P}_k &= \frac{P_oR}{2kP_0 + R} \\ \bar{S}_k &= \frac{2kP_oR + R^2}{2[(k-1)P_0 + R]} \end{aligned} \tag{B.4}$$

Equations (B.5) and (B.6) represent the Kalman gains as a function of initial values and with different measurement noise variances.

$$K_k = \frac{P_{k-1}}{P_{k-1} + R} = \frac{P_0}{kP_0 + R} \tag{B.5}$$

$$\bar{K}_k = \frac{\bar{P}_{k-1}}{\bar{P}_{k-1} + R} = \frac{2P_0}{2kP_0 + R} \tag{B.6}$$

Now, we write $\hat{x}_{k|k}$ and $\bar{x}_{k|k}$ as function of time index k for parameter estimation case when $H = 1$.

$$\hat{x}_{1|1} = x_0 + K_1(y_1 - x_0)$$

Putting the value of K_1 from equation (B.5)

$$\begin{aligned} \hat{x}_{1|1} &= x_0 + \frac{P_0}{P_0 + R}(y_1 - x_0) \\ &= x_0 + \frac{P_0R}{P_0 + R}y_1 - \frac{P_0}{P_0 + R}y_1x_0 \\ &= \frac{R}{P_0 + R}y_1x_0 + \frac{P_0}{P_0 + R}y_1 \\ &= \frac{1}{P_0 + R}(Rx_0 + P_0y_1) \end{aligned}$$

Now, we find expression of $\hat{x}_{2|2}$

$$\hat{x}_{2|2} = \hat{x}_{1|1} + K_2(y_2 - \hat{x}_{1|1})$$

Putting the value of K_2 from equation (B.5)

$$\begin{aligned}\hat{x}_{2|2} &= \hat{x}_{1|1} + \frac{P_0}{2P_0 + R}(y_2 - \hat{x}_{1|1}) \\ &= \hat{x}_{1|1} - \frac{P_0}{2P_0 + R}\hat{x}_{1|1} + \frac{P_0}{2P_0 + R}y_2 \\ &= \frac{P_0 + R}{2P_0 + R}\hat{x}_{1|1} + \frac{P_0}{2P_0 + R}y_2\end{aligned}$$

Putting the value of $\hat{x}_{1|1}$ from above equation

$$\begin{aligned}\hat{x}_{2|2} &= \frac{P_0 + R}{2P_0 + R}(\frac{1}{P_0 + R}(Rx_0 + P_0y_1)) + \frac{P_0}{2P_0 + R}y_2 \\ &= \frac{R}{2P_0 + R}x_0 + \frac{P_0}{2P_0 + R}(y_1 + y_2)\end{aligned}$$

Now, we derive the expression for $\hat{x}_{3|3}$ in below set of equations.

$$\hat{x}_{3|3} = \hat{x}_{2|2} + K_3(y_3 - \hat{x}_{2|2})$$

Putting the value of K_3 from equation (B.5)

$$\begin{aligned}\hat{x}_{3|3} &= \hat{x}_{2|2} + \frac{P_0}{3P_0 + R}(y_3 - \hat{x}_{2|2}) \\ &= \hat{x}_{2|2} - \frac{P_0}{3P_0 + R}\hat{x}_{2|2} + \frac{P_0}{3P_0 + R}y_3 \\ &= \frac{2P_0 + R}{3P_0 + R}\hat{x}_{2|2} + \frac{P_0}{3P_0 + R}y_3\end{aligned}$$

Putting the value of $\hat{x}_{2|2}$ from above equation

$$\begin{aligned}\hat{x}_{3|3} &= \frac{2P_0 + R}{3P_0 + R}\left(\frac{R}{2P_0 + R}x_0 + \frac{P_0}{2P_0 + R}(y_1 + y_2)\right) + \frac{P_0}{3P_0 + R}y_3 \\ &= \frac{R}{3P_0 + R}x_0 + \frac{P_0}{3P_0 + R}(y_1 + y_2 + y_3)\end{aligned}$$

Using the induction procedure, we can write the expression of $\hat{x}_{k|k}$ in terms of k, x_0 and P_0 as below:

$$\hat{x}_{k|k} = \frac{R}{kP_0 + R} x_0 + \frac{P_0}{kP_0 + R} \sum_{i=1}^{k} y_i \tag{B.7}$$

Now, we write the expression of $\bar{x}_{k|k}$ in terms of k, x_0 and P_0. We write the general expression using an induction as before and using the Kalman gain expressed in equation (B.6).

$$\bar{x}_{k|k} = \frac{R}{2kP_0 + R} x_0 + \frac{2P_0}{2kP_0 + R} \sum_{i=1}^{k} y_i \tag{B.8}$$

B.3 Reduction of $L(m,n)$ (Section 6.3.1)

We work on the expression of $\sum_{i=m+1}^{n} \hat{J}_i$ to write it in reduced form.

$$\begin{aligned}\sum_{i=m+1}^{n} \hat{J}_i = \log\left(\frac{(m+1)P_o + R}{mP_0 + R}\right) + \log\left(\frac{(m+2)P_o + R}{(m+1)P_0 + R}\right) + ... + \log\left(\frac{(n-1)P_o + R}{(n-2)P_0 + R}\right) \\ + \log\left(\frac{nP_o + R}{(n-1)P_0 + R}\right)\end{aligned}$$

Using the basic formula of $\log(A/B) = \log(A) - \log(B)$, we get below set of expressions.

$$\begin{aligned}\sum_{i=m+1}^{n} \hat{J}_i = \log((m+1)P_o + R) - \log(mP_0 + R) + \log((m+2)P_o + R) - \log((m+1)P_0 + R) \\ + ... + \log((n-1)P_o + R) - \log((n-2)P_0 + R) + \log(nP_o + R) - \log((n-1)P_0 + R)\end{aligned}$$

Cancelling the same terms we get:

$$\begin{aligned}\sum_{i=m+1}^{n} \hat{J}_i &= -\log(mP_0 + R) + \log(nP_o + R) \\ &= \log\left(\frac{nP_o + R}{mP_0 + R}\right)\end{aligned} \tag{B.9}$$

Similarly, we can calculate $\sum_{i=m+1}^{n} \bar{J}_i$ using above induction procedure.

$$\sum_{i=m+1}^{n} \bar{J}_i = \log\left(\frac{2(m+1)P_o + R}{2mP_0 + R}\right) + \log\left(\frac{2(m+2)P_o + R}{2(m+1)P_0 + R}\right) + ... + \log\left(\frac{2(n-1)P_o + R}{2(n-2)P_0 + R}\right) + \log\left(\frac{2nP_o + R}{2(n-1)P_0 + R}\right)$$

Using the basic formula of $\log(A/B) = \log(A) - \log(B)$, we get below set of expressions.

$$\begin{aligned}\sum_{i=m+1}^{n} \bar{J}_i = {} & \log(2(m+1)P_o + R) - \log(2mP_0 + R) + \log(2(m+2)P_o + R) \\ & - \log(2(m+1)P_0 + R) + ... + \log(2(n-1)P_o + R) - \log(2(n-2)P_0 + R) \\ & + \log(2nP_o + R) - \log(2(n-1)P_0 + R)\end{aligned}$$

Cancelling the same terms we get:

$$\begin{aligned}\sum_{i=m+1}^{n} \bar{J}_i &= -\log(2mP_0 + R) + \log(2nP_o + R) \\ &= \log\left(\frac{2nP_o + R}{2mP_0 + R}\right)\end{aligned} \tag{B.10}$$

B.4 Divergence Test of Sequence a_k

In this Appendix, we start looking on the divergence of the sequence a_k which is defined as partial sums $a_k = \sum_{i=1}^{k} b_i = \log(T_k)$

If the value of $b_k = 0$ when $k \to \infty$ then the series a_k is not divergent but it does not tell about its convergence. In order to prove its convergence we need more information.

$$\begin{aligned}b_k &= \frac{1}{2}\left(J_k + \hat{Q}_k - \bar{Q}_k\right) \\ \lim_{k\to\infty} b_k &= \lim_{k\to\infty} \frac{1}{2}\left(J_k + \hat{Q}_k - \bar{Q}_k\right)\end{aligned} \tag{B.11}$$

We evaluate the limiting value of J_k when $k = \infty$ first as below.

$$\begin{aligned} J_{k+1} &= 2\log\left(\frac{kP_0 + R}{(k-1)P_0 + R}\right) - \log\left(\frac{2kP_0 + R}{2(k-1)P_0 + R}\right) \\ &= 2\log\left(\frac{kP_0 + R}{kP_0 - P_0 + R}\right) - \log\left(\frac{2kP_0 + R}{2kP_0 - 2P_0 + R}\right) \\ &= 2\log\left(\frac{P_0 + R/k}{P_0 - P_0/k + R/k}\right) - \log\left(\frac{P_0 + R/2k}{P_0 - 2P_0/2k + R/2k}\right) \end{aligned}$$

Evaluating the above expression by putting $k \to \infty$ in above simplified expression

$$\begin{aligned} \lim_{k\to\infty} J_k &= 2\log\left(\frac{P_0 + R/\infty}{P_0 - P_0/\infty + R/\infty}\right) - \log\left(\frac{P_0 + R/\infty}{P_0 - 2P_0/\infty + R/\infty}\right) \\ &= 2\log\left(P_0/P_0\right) - \log\left(P_0/P_0\right) \\ &= 2\log(1) - \log(1) \\ &= 0 \end{aligned}$$

The limiting value of the term J_k approaches to zero. Now, we can evaluate the limiting values of $\bar{Q}_k$ and $\hat{x}_k$ when $k \to \infty$.

$$\begin{aligned} \lim_{k\to\infty} \bar{Q}_k &= \lim_{k\to\infty} \frac{2}{R}(y_k - H\bar{x}_k)^2 \\ \lim_{k\to\infty} \hat{Q}_k &= \frac{2}{R}(y_k - H\hat{x}_k)^2 \end{aligned}$$

As we have proved earlier in lemma 6.2 that for large $k \to \infty$, the value of $\bar{x}_{k|k}$ is equal to value of $\hat{x}_{k|k}$ therefore observing above equation, it is clear that $\lim_{k\to\infty} \bar{Q}_k = \lim_{k\to\infty} \hat{Q}_k$.

$$\begin{aligned} \lim_{k\to\infty} b_\infty &= \lim_{k\to\infty} \frac{1}{2}(J_k + \hat{Q}_k - \bar{Q}_k) \\ &= \frac{1}{2}(\lim_{k\to\infty} J_k + \lim_{k\to\infty} \hat{Q}_k - \lim_{k\to\infty} \bar{Q}_k) \end{aligned} \tag{B.12}$$

Using the above worked results, we can write the limiting value of b_k as below.

$$\lim_{k \to \infty} b_\infty \to 0 \tag{B.13}$$

We have shown that when $k \to \infty$ the value of b_k also approaches to 0, therefore the partial sum sequence a_k is not divergent.

Bibliography

[1] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188, 2002.

[2] Yaakov Bar-Shalom, X Rong Li, and Thiagalingam Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software*. Wiley-Interscience, 2001.

[3] Yaakov Bar-Shalom and Xiao-Rong Li. Multitarget-multisensor tracking: principles and techniques. *Storrs, CT: University of Connecticut, 1995.*, 1995.

[4] Yaakov Bar-Shalom and Edison Tse. Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11(5):451–460, 1975.

[5] Yaakov Bar-Shalom and Edison Tse. Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11(5):451–460, 1975.

[6] B Belkin, SL Anderson, and KM Sommar. The pseudo-measurement approach to track-to-track data fusion. In *Proc. 1993 Joint Service Data Fusion Symposium*, pages 519–538, 1993.

[7] Samuel S Blackman. Multiple-target tracking with radar applications. *Dedham, MA, Artech House, Inc., 1986, 463 p.*, 1, 1986.

[8] Samuel S Blackman and Robert Popoli. *Design and analysis of modern tracking systems*, volume 685. Artech House Norwood, MA, 1999.

[9] Richard S Bucy and Kenneth D Senne. Digital synthesis of non-linear filters. *Automatica*, 7(3):287–298, 1971.

[10] Aniello Buonocore, Enrica Pirozzi, and Luigia Caputo. A note on the sum of uniform random variables. *Statistics & Probability Letters*, 79(19):2092–2097, 2009.

[11] Czesław Bylinski. The sum and product of finite sequences of real numbers. *Journal of Formalized Mathematics*, 2, 1990.

[12] Olivier Cappé, Simon J Godsill, and Eric Moulines. An overview of existing methods and recent advances in sequential monte carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007.

[13] D Castanon and Demosthenis Teneketzis. Distributed estimation algorithms for nonlinear systems. *Automatic Control, IEEE Transactions on*, 30(5):418–425, 1985.

[14] Subhash Challa, Mark R Morelande, Darko Mušicki, and Robin J Evans. *Fundamentals of object tracking*. Cambridge University Press, 2011.

[15] Subhash Challa, Mark R Morelande, Darko Mušicki, and Robin J Evans. *Fundamentals of object tracking*. Cambridge University Press, 2011.

[16] Zhe Chen. Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69, 2003.

[17] J Cornebise, E Moulines, and J Olsson. Adaptive methods for sequential importance sampling. *Journ ées MAS de la SMAI, Rennes, France*, 2008.

[18] Thomas M Cover and Joy A Thomas. *Elements of information theory*. Wiley-interscience, 2006.

[19] James Davidson. *Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians*. Oxford university press, 1994.

[20] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

[21] Oliver E Drummond. Feedback in track fusion without process noise. In *SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation*, pages 369–383. International Society for Optics and Photonics, 1995.

[22] Oliver E Drummond. Hybrid sensor fusion algorithm architecture and tracklets. In *Optical Science, Engineering and Instrumentation'97*, pages 485–502. International Society for Optics and Photonics, 1997.

[23] Vassiliy A Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158, 1969.

[24] Gabriel Frenkel. Multisensor tracking of ballistic targets. In *SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation*, pages 337–346. International Society for Optics and Photonics, 1995.

[25] LA Gerhardt and Kenneth W Drake. Estimation of the shape of probability density functions using deterministic algorithms. In *Decision and Control, 1971 IEEE Conference on*, volume 10, pages 337–341. IEEE, 1971.

[26] Mark Girolami and Chao He. Probability density estimation from optimally condensed data samples. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1253–1264, 2003.

[27] Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET, 1993.

[28] JE Handschin. Monte carlo techniques for prediction and filtering of non-linear stochastic processes. *Automatica*, 6(4):555–563, 1970.

[29] JE Handschin and David Q Mayne. Monte carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering? *International journal of control*, 9(5):547–559, 1969.

[30] Simon S Haykin, Simon S Haykin, and Simon S Haykin. *Kalman filtering and neural networks*. Wiley Online Library, 2001.

[31] F van der Heijden. Consistency checks for particle filters. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):140–145, 2006.

[32] Yu-Chi Ho and RCKA Lee. A bayesian approach to problems in stochastic estimation and control. *Automatic Control, IEEE Transactions on*, 9(4):333–339, 1964.

[33] Jenq-Neng Hwang, S-R Lay, and Alan Lippman. Nonparametric multivariate density estimation: a comparative study. *Signal Processing, IEEE Transactions on*, 42(10):2795–2810, 1994.

[34] Bo Jiang and Binoy Ravindran. Completely distributed particle filters for target tracking in sensor networks. In *Parallel & Distributed Processing Symposium (IPDPS), 2011 IEEE International*, pages 334–344. IEEE, 2011.

[35] Simon J Julier and Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics, 1997.

[36] T Kailath, AH Sayed, and B Hassibi. Linear estimation prentice-hall. *Upper Saddle River, NJ*, 2000.

[37] Rudolph E Kalman and Richard S Bucy. New results in linear filtering and prediction theory. *Journal of Fluids Engineering*, 83(1):95–108, 1961.

[38] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

[39] Thiagalingam Kirubarajan and Yaakov Bar-Shalom. Probabilistic data association techniques for target tracking in clutter. *Proceedings of the IEEE*, 92(3):536–557, 2004.

[40] Barbara F La Scala and Robin J Evans. Minimum necessary data rates for accurate track fusion. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 6966–6971. IEEE, 2005.

[41] X Rong Li, Yunmin Zhu, Jie Wang, and Chongzhao Han. Optimal linear estimation fusion. i. unified fusion rules. *Information Theory, IEEE Transactions on*, 49(9):2192–2208, 2003.

[42] John MacCormick and Andrew Blake. A probabilistic exclusion principle for tracking multiple objects. *International Journal of Computer Vision*, 39(1):57–71, 2000.

[43] Jean-Michel Marin, Pierre Pudlo, and Mohammed Sedki. Consistency of the adaptive multiple importance sampling. *arXiv preprint arXiv:1211.2548*, 2012.

[44] D Musicki and Robin J Evans. Track fusion using equivalent innovations. In *Information Fusion, 2007 10th International Conference on*, pages 1–8. IEEE, 2007.

[45] Darko Musicki and Robin Evans. Joint integrated probabilistic data association: Jipda. *Aerospace and Electronic Systems, IEEE Transactions on*, 40(3):1093–1099, 2004.

[46] Darko Musicki, Robin Evans, and Srdjan Stankovic. Integrated probabilistic data association. *Automatic Control, IEEE Transactions on*, 39(6):1237–1241, 1994.

[47] Darko Musicki and Robin J Evans. Multiscan multitarget tracking in clutter with integrated track splitting filter. *Aerospace and Electronic Systems, IEEE Transactions on*, 45(4):1432–1447, 2009.

[48] Darko Musicki, Barbara F La Scala, and Robin J Evans. Integrated track splitting filter-efficient multi-scan single target tracking in clutter. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(4):1409–1425, 2007.

[49] Darko Musicki and Taek Lyul Song. Track initialization: Prior target velocity and acceleration moments. *Aerospace and Electronic Systems, IEEE Transactions on*, 49(1):665–670, 2013.

[50] Nickens N Okello and Subhash Challa. Joint sensor registration and track-to-track fusion for distributed trackers. *Aerospace and Electronic Systems, IEEE Transactions on*, 40(3):808–823, 2004.

[51] Yaakov Oshman. Optimal sensor selection strategy for discrete-time state estimators. *Aerospace and Electronic Systems, IEEE Transactions on*, 30(2):307–314, 1994.

[52] Francesco Palmieri, Stefan Marano, and Peter Willett. Measurement fusion for target tracking under bandwidth constraints. In *Aerospace Conference, 2001, IEEE Proceedings.*, volume 5, pages 2179–2190. IEEE, 2001.

[53] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, pages 1065–1076, 1962.

[54] Michael K Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.

[55] Muhammad A Raza, Mark R Morelande, and Robin J Evans. Distributed target tracking under communication constraints. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 4035–4039. IEEE, 2015.

[56] Fazlollah M Reza. *An introduction to information theory*. Dover Publications, 2010.

[57] Robert and W.P Novinger. *Complex Variables*, volume 1. http://www.math.uiuc.edu/ r-ash/CV.html, 1971.

[58] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2004.

[59] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[60] JA Roecker. Multiple scan joint probabilistic data association. *Aerospace and Electronic Systems, IEEE Transactions on*, 31(3):1204–1210, 1995.

[61] Len J Sciacca and Robin J Evans. Cooperative sensor networks with bandwidth constraints. In *SPIE proceedings series*, pages 192–201. Society of Photo-Optical Instrumentation Engineers, 2002.

[62] Y Bar Shalom and TE Fortmann. Tracking and data association. *Number 179 in Mathematics in science and engineering. Academic Press, Boston*, 1988.

[63] M Simandl and Ondrej Straka. Sampling densities of particle filter: a survey and comparison. In *American Control Conference, 2007. ACC'07*, pages 4437–4442. IEEE, 2007.

[64] Robert W Sittler. An optimal data association problem in surveillance theory. *Military Electronics, IEEE Transactions on*, 8(2):125–139, 1964.

[65] E Skafidas, RJ Evans, and A Logothetis. Data fusion by optimal sensor switching. In *Data Fusion Symposium, 1996. ADFS'96., First Australian*, pages 111–113. IEEE, 1996.

[66] Václav Smídl and Radek Hofman. Adaptive importance sampling in particle filtering. In *Information Fusion (FUSION), 2013 16th International Conference on*, pages 9–16. IEEE, 2013.

[67] Taek Song, Hyoung Kim, and Darko Musicki. Distributed (nonlinear) target tracking in clutter. *Aerospace and Electronic Systems, IEEE Transactions on*, 51(1):654–668, 2015.

[68] Harold W SORENSON and STUBBERUD. Non-linear filtering by approximation of the a posteriori density.

[69] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. IEEE, 2000.

[70] Roy D Yates and David J Goodman. Probability and stochastic processes. *John Willey & Sons*, 1999.

[71] RD Yatess and DJ Goodman. Probability and stochastic processes. a friendly introduction for electrical and computer engineering. 2005.

[72] Keshu Zhang, X Rong Li, Peng Zhang, and Haifeng Li. Optimal linear estimation fusionpart vi: Sensor data compression. In *Proc. Int. Conf. Information Fusion*, volume 23, page 221, 2003.

[73] Yunmin Zhu, Enbin Song, Jie Zhou, and Zhisheng You. Optimal sensor data linear compression in multisensor estimation fusion. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 6, pages 5807 – 5812 Vol.6, dec. 2003.