Contributions to filtering under randomly delayed observations and additive-multiplicative noise

Seham Allahyani

A thesis submitted for the degree of Doctor of Philosophy

Department of Mathematics
College of Engineering, Design and Physical Sciences

Brunel University London

2017

Abstract

This thesis deals with the estimation of unobserved variables or states from a time series of noisy observations. Approximate minimum variance filters for a class of discrete time systems with both additive and multiplicative noise, where the measurement might be delayed randomly by one or more sample times, are investigated. The delayed observations are modelled by up to N sample times by using N Bernoulli random variables with values of 0 or 1. We seek to minimize variance over a class of filters which are linear in the current measurement (although potentially nonlinear in past measurements) and present a closed-form solution. An interpretation of the multiplicative noise in both transition and measurement equations in terms of filtering under additive noise and stochastic perturbations in the parameters of the state space system is also provided. This filtering algorithm extends to the case when the system has continuous time state dynamics and discrete time state measurements. The Euler scheme is used to transform the process into a discrete time state space system in which the state dynamics have a smaller sampling time than the measurement sampling time. The number of sample times by which the observation is delayed is considered to be uncertain and a fraction of the measurement sample time.

The same problem is considered for nonlinear state space models of discrete time systems, where the measurement might be delayed randomly by one sample time. The linearisation error is modelled as an additional source of noise which is multiplicative in nature. The algorithms developed are demonstrated throughout with simulated examples.

Acknowledgements

In the name of Allah, the Most Gracious, the Most Merciful

I would like to take this opportunity to thank those who help me complete this thesis with encouragement, support and inspiration they afford to me. First and foremost, my sincere gratitude goes to, my supervisor, Dr. Paresh Date. Dr. Paresh consistently supported me to accomplish this PhD thesis. During my PhD journey, Dr. Paresh was an exceptional guide and great supervisor. I say, to Dr. Paresh, thanks a lot for the support and assistance to make this work achievable.

I am grateful to my great mother, Kamelah. I would like to thank her for support and prayers. In addition, I would like to thanks my husband Faris and my two kids Abdullah and Sattam who supported me to complete this PhD. Your care, help and love are appreciated. To them I say Im beholden and your support was the force behind my success. Finally, I am thankful to my sisters and brothers.

Contents

Notation

The notation used throughout the chapters is introduced here. 'time' refers to sampling time instant in this table and 'rv' stands for 'random variable'.

$\mathcal{X}(k)$	State vector at time k
$\mathcal{Y}(k)$	Measurement vector at time k
A, C, G_1, G_2, U_w, U_v	Constant matrices
B, D	Constant vectors
$\mathcal{W}(k), \mathcal{V}(k)$	Vector-valued rvs representing additive noise at time k
$\mathcal{W}(t)$	Standard Wiener process
$\mathcal{S}_1(k), \mathcal{S}_2(k)$	Vector-valued rvs representing multiplicative noise at time k
p_k	Bernoulli random variable representing a random delay at time k
$A^\top$	Transpose of a vector or matrix A
$\mathbb{P}(A)$	Probability of an event A
$\mathbb{E}[\mathcal{X}]$	Expected value of a rv $\mathcal{X}$
f, h	nonlinear, vector-valued functions
$tr(\cdot)$	The trace of a matrix
$(\star)^\top$	Transpose of a matrix-valued expression
$diag(y_k)$	A block diagonal matrix with y_k as the k^{th} diagonal element
$vec(f(x_j))$	A vector with $f(x_j)$ as its j^{th} element
$\hat{\mathcal{X}}(k+i\|k)$	Estimate of $\mathcal{X}(k+i), i \geq 0$ from $\mathcal{Y}(k-j), j = 0, 1, \cdots$

Acronyms

AvRMSE	The average of RMSE
CDF	Continuous-discrete filtering
DDF	Discrete-discrete filtering
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
GRV	Gaussian random variables
MKF	Modified Kalman Filter
MV	Minimum variance
ODE	Ordinary differential equation
PDF	Probability density function
PF	Particle Filter
QKF	Quadrature Kalman Filter
RMSE	Root mean square error
SDE	Stochastic differential equation
UKF	Unscented Kalman Filter

List of Tables

Chapter 1

Introduction

This introductory chapter presents the motivation for studying certain problems within the field of latent state estimation and describes the main contributions of the thesis. It also outlines the structure of the rest of the thesis.

1.1 Motivation

In several branches of computer science and engineering applications, we need to estimate unobserved variables or states from a time series of noisy observations, given a functional relationship between them. A commonly used technique to deal with this problem is called a *filter*. Research on filtering and estimation problems has attracted a great deal of attention due to their extensive applications in many practical areas, including economics, tracking, weather forecasting, navigation systems, control and signal processing. The central idea behind the filtering problems is building an approximation of the posterior probability density of the unobserved state variables or model parameters, recursively in time, by incorporating new noisy observations and model predictions. For any system, the complete information about the state can be found in the probability density functions (PDFs) of the state. In general, the solution of the filtering problem is based on Bayes' theorem, which tells us how to update the PDF of the unobserved state, the so-called *prior* PDF, with new observations to obtain the so-called *posterior* PDF. For discrete time systems with a constant sampling time, such a filter works as a two-step process. First, the prior PDF at time k is obtained using the known dynamics of the state and the posterior PDF at time k-1. Once the measurement at time k arrives, this PDF is updated using the likelihood of the observed measurement.

One of the most common approaches to latent state estimation in linear dynamical systems is the Kalman filter, which provides an optimal solution in the sense of minimum variance (MV) for linear systems with Gaussian additive noise. However, in most applications of interest, the state transition and the observation equations are nonlinear and the Kalman filter cannot be directly applied. As a result, a large variety of approximate nonlinear filters have been considered, such as the extended Kalman filter (EKF), the particle filter (PF), the unscented Kalman filter (UKF), the Ensemble Kalman filter (EnKF) and the quadrature Kalman filter (QKF). Some of these filtering algorithms will be examined in more detail in the chapter two.

Most traditional filter design approaches depend on the assumption that measurement signals are perfectly transmitted. However, in many practical situations the measurements available in estimation do not arrive in time but are randomly delayed due to several factors like slow sensors, long processing time of the sensor data, limited capacity of the communication link, radar-based devices, *etc.* Further, it is worth pointing out that most traditional filtering algorithms are only concerned with the additive noises. However, systems may also contain multiplicative noise, which may act as a proxy for some classes of parameter uncertainties or for linearisation error. It is known that, for a linear system with a multiplicative noise, the optimal filter is nonlinear.

1.2 Contribution

In this thesis several filtering algorithms which tackle the issues of multiplicative noise and measurement delayed are developed. The thesis discusses both discrete and continuous-discrete time filtering problems. The main contributions of the thesis are as follows:

1. An approximate minimum variance (MV) filter for a class of discrete and continuous-discrete time linear systems with both additive and multiplicative noise is investigated. An interpretation of the multiplicative noise in terms of parameter perturbations in a linear additive model is provided. For continuous-discrete time systems, the Euler scheme followed by conditional moment matching to transform stochastic differential equations (SDEs) in the process equation into a discrete model on a timescale which is finer than the measurement timescale is used.

2. A new state estimation algorithm for unobserved state variables is proposed for a class of discrete and continuous-discrete time linear systems with both additive and multiplicative noise, where the measurement might be delayed randomly by one or more sample times. Specifically, a filter for this class of systems is constructed which minimizes the trace of the estimation error covariance matrix over a set of filters which are linear in the current measurement while being nonlinear in one or more past measurements. The number of sample times by which the observation is delayed is considered to be uncertain. For continuous-discrete time systems, the Euler scheme is used to transform the process into a discrete time state space system where the state dynamics have a smaller sampling time than the measurement sampling time.

3. An approximate MV filter for discrete time nonlinear systems with randomly delayed observations and additive and multiplicative noise is proposed. The model is linearized at each time step, and the linearisation error is modelled in terms of multiplicative stochastic noise in

the linearized state space model. A heuristic is provided to adjust the size of linearisation error in terms of the trace of the estimation error covariance matrix, in order to tune the filter. Once the nonlinear system is reduced to a discrete time linearised system with both additive and multiplicative noise, one can use the tools developed earlier in the thesis to design an approximate MV filter in the presence of random delays.

1.3 Thesis Structure

The rest of the thesis is structured as follows.

- In Chapter 2, a brief overview of state estimation and filtering problems is provided. A brief description and review of some filtering algorithms is also provided. In particular, we briefly describe the linear Kalman filter and its algorithm in some detail. Some nonlinear filter algorithms, particularly the EKF and the UKF, are explored in some detail, and their advantages and disadvantages are discussed. Moreover, a brief review of particle filtering is presented, along with a numerical example.

- In Chapter 3, an interpretation of filtering in additive noise models under parameter perturbations in some of the parameters is considered. A multiplicative noise term is considered in both the process and the measurement equations to deal with the stochastic uncertainty that arises out of either linearisation errors or parametric uncertainty. The performance of the new filter is compared to the existing result. In addition, a new algorithm for approximate MV filtering in the presence of both additive and multiplicative noise, when the measurements are randomly delayed and the delay is an integer number of time steps, is provided. The filter is linear in the current measurement while it was nonlinear in past measurements. A complete closed-form solution to the delayed filtering problem for this class of systems is proposed.

- In chapter 4, the approximate MV filter proposed in chapter 3 for discrete state space systems with multiplicative noise is extended to continuous-discrete systems with multiplicative noise. The differential equations that describe the process are discretised using the Euler scheme at a higher sampling frequency than the measurement frequency. The procedure for deriving the expressions of a filter with randomly delayed measurements where the delay is a fraction of a time step is outlined.

- In chapter 5, an approximate MV filter for discrete time nonlinear systems with randomly delayed observations and additive and multiplicative noise is investigated. Specifically, an interpretation of filtering under multiplicative noise in terms of linearisation error is provided, and a closed-form MV filter for this system is derived. Its performance is compared with the EKF.

- In chapter 6, the thesis concludes with a summary of the main contributions of the thesis and suggestions for future work.

1.4 Publications

- S. Allahyani, P. Date, An Approximate Minimum Variance Filter for Nonlinear Systems with Randomly Delayed Observations, in: 25th European Signal Processing Conference (EUSIPCO 2017), IEEE, 2017, pp. 1639-1643

- S. Allahyani, P. Date, A minimum variance filter for continuous discrete systems with additive-multiplicative noise, in: 24th European Signal Processing Conference (EUSIPCO 2016), IEEE, 2016, pp. 2330-2334.

- S. Allahyani, P. Date, A minimum variance filter for discrete time linear systems with parametric uncertainty, in: MED'16: The 24th Mediterranean Conference on Control and Automation, Mediterranean Control

Association, 2016, pp. 159-163.

- S. Allahyani, P. Date, A new approximate minimum variance filter for discrete time linear systems with randomly delayed observations and additive-multiplicative noise, Under revision.
- S. Allahyani, P. Date, A new approximate minimum variance filter for continuous-discrete linear systems with randomly delayed observations and additive-multiplicative noise, Submitted for journal publication.

Chapter 2

Preliminaries

In this chapter the state estimation is introduced and the filtering problem is described. The Kalman filter [1] and its limitations are reviewed before several algorithms for approximate nonlinear filtering are described in detail.

2.1 State Estimation

The state estimation in stochastic state space models can be described in terms of finding probabilistic information about a *state* vector $\mathcal{X}(k)$ which typically evolves over time. The state vector $\mathcal{X}(k)$ itself is in general not directly observable. Instead, measurements vector $\mathcal{Y}(k)$, which are a noisy function in the current state vector $\mathcal{X}(k)$, are observed at specific times. In many application areas, the model state is represented by an (approximated) PDF. State estimation is a means to propagate the PDF of the system states over time in some optimal way. It is most common to use the Gaussian PDF to represent the model state, process and measurement noises. The user, however, is in general interested in recovering the PDF of $\mathcal{X}(k)$ given a stream of measurements online up to time l, $\mathcal{Y}(1:l) = \{\mathcal{Y}(1), \ldots, \mathcal{Y}(l)\}$, in order to calculate up-to-date estimates of the system states. The interpretation of the state as a random variable with a certain distribution is in accordance with the Bayesian viewpoint, and the sought after conditional density $p(\mathcal{X}(k)|\mathcal{Y}(1:k))$ is often called the posterior density, where $\mathcal{Y}(1:k))$ are measurements vector up to time k. For any variable or parameter in estimation, there are three quantities of interest: the true value, the measured value and the estimated value. The true value (or truth) is usually unknown in practice; this represents the actual value sought of the quantity being approximated by the estimator. The measured value denotes the quantity which is directly determined from a sensor. Measurements are never perfect, since they will always contain errors. Thus, measurements are usually modeled using a function of the true values plus some error. The estimated values of the latent state are found using a combination of a dynamic state transition model and the measurement model.

Based on the relation between the time of interest k and the time interval of available observations, determined by l, three different estimation problems are recognized, due to the different relations between k and l filtering, smoothing and prediction.

Filtering: In a Bayesian filtering problem $l = k$ (i.e., observations including frame l), that is all the current and past observations are available to estimate the state. Clearly, the basis of the proposed filtering approach is to develop an online algorithm to compute the filtering density $p(\mathcal{X}(k)|\mathcal{Y}(1:k))$ when a measurement $\mathcal{Y}(k)$ becomes available.

Smoothing: In a Bayesian smoothing problem $l > k$, that is both past and future measurements are available to compute the filtering density $p(\mathcal{X}(k)|\mathcal{Y}(1:l))$.

Prediction: In a Bayesian prediction problem $l < k$, that is the available measurements $\mathcal{Y}(1:l)$ are used to compute a prediction density of the state in the future.

2.2 The Filtering Problem

Based on the available information (control inputs, if present, and observation), it is required to obtain an estimate of the system's state that optimizes a given criteria. This is the role played by a *filter*. Filtering approaches estimate the unknown true state $\mathcal{X}(k)$ from certain noisy observations $\mathcal{Y}(k)$. Within the Bayesian framework, we specified the filtering problem in terms of finding a conditional probability density $p(\mathcal{X}(k)|\mathcal{Y}(1:k))$, the filtering posterior. The user, however, is in general interested in a point estimate $\hat{\mathcal{X}}(k|k)$ of the state and some indication of its confidence, often represented by a covariance matrix $\bar{P}(k|k)$. The predominant idea behind point estimate computations is the minimization of some expected loss, where the expectation is carried out with respect to the posterior density. An interesting discussion can be found in [2], where it is shown that the conditional mean $E[\mathcal{X}(k)|\mathcal{Y}(1:k)]$ minimizes the expected squared estimation error regardless of the shape of $p(\mathcal{X}(k)|\mathcal{Y}(1:k))$. The conditional mean is therefore also known as the minimum mean squared error or MMSE estimate.

The general filtering problem may the formulated as follows [2]:

$$\mathcal{X}(k+1) = f(\mathcal{X}(k), \mathcal{U}(k), \mathcal{W}(k)), \tag{2.1}$$

$$\mathcal{Y}(k) = h(\mathcal{X}(k), \mathcal{V}(k)). \tag{2.2}$$

The state difference equation (2.1) relates the upcoming state $\mathcal{X}(k+1)$ to the current state $\mathcal{X}(k)$ and the process noise $\mathcal{W}(k)$, via a function f. The measurement equation (2.2) relates $\mathcal{X}(k)$ and the measurement noise $\mathcal{V}(k)$ to the observation $\mathcal{Y}(k)$, via a function h. All of the quantities $\mathcal{X}(k)$, $\mathcal{W}(k)$, $\mathcal{Y}(k)$ and $\mathcal{V}(k)$ are vectors with known dimensions. $\mathcal{U}(k) \in R^n$ is the control input. In control engineering, there is often a separate exogenous input term in equation (2.1); this term is not considered here since the focus is simply on filtering rather than control design and hence the input term is not relevant. The time increment $t(k) - t(k-1)$ is assumed to be constant for all k. The following assumptions are made for the problems that we treat in this thesis. The initial time index is $k = 0$ and the first measurement is taken at $k = 1$. Both functions f and h are known for each time instance k. The noise characteristics of $\mathcal{W}(k)$ and $\mathcal{V}(k)$ are known, as is the distribution of $\mathcal{X}(0)$. We assume $\mathcal{W}(k)$ and $\mathcal{V}(k)$ to be uncorrelated, such that is $cov\{\mathcal{W}(k), \mathcal{W}(l)\} = 0$ and $covariance\{\mathcal{V}(k), \mathcal{V}(l)\} = 0$ for $k \neq l$. The initial state $\mathcal{X}(0)$ is assumed to be uncorrelated to $\mathcal{W}(k)$ and $\mathcal{V}(k)$. The filtering problem is the problem of finding the conditional distribution $\mathbb{P}(\mathcal{X}(k)|\mathcal{X}(k-1), \mathcal{Y}(k))$. This takes the form of finding a discrete approximation of the distribution, or of finding the first two conditional moments of the distribution. An exact solution to the filtering problem is possible if f and h are linear in the underlying variables and all the noise sources are Gaussian. This solution is described later in section 2.3.1.

2.2.1 The Discrete Time Linear System

The linear state space model that is most common in the literature and has been best researched and understood is as follows:

$$\mathcal{X}(k+1) = A\mathcal{X}(k) + B + U_w\mathcal{W}(k), \tag{2.3}$$

$$\mathcal{Y}(k) = C\mathcal{X}(k) + D + U_v\mathcal{V}(k), \tag{2.4}$$

which has been at the core of estimation theory for several decades. The variable $\mathcal{X}(k) \in \mathbb{R}^n$ is the state vector at time k and needs to be estimated, $\mathcal{Y}(k) \in \mathbb{R}^r$ is the measurement vector at time k, A, C, U_w and U_v are given constant matrices and B and D are given constant vectors of compatible dimensions. Note that equations (2.3)-(2.4) are a special case of system (2.1)-(2.2) with linear functions. For linear Gaussian systems (2.3)-(2.4), the filtering density is Gaussian throughout time and as such fully described by an estimate $\hat{\mathcal{X}}(k|k)$ and its covariance $\bar{P}(k|k)$. In fact, $\hat{\mathcal{X}}(k|k)$ and $\bar{P}(k|k)$ are what are called sufficient statistics for the Gaussian distribution. There exists an optimal unbiased state estimation algorithm that provides the lowest mean squared estimation error: the celebrated Kalman filter that we discuss in Section 2.3.1. It should be noted that (2.3)-(2.4) could be obtained by linearisation of (2.1)-(2.2), an idea which is exploited in a family of approximate filtering algorithms, including the EKF, in Section 2.3.2.

2.2.2 The Discrete Time Nonlinear System

A widely used special case of (2.1)-(2.2) is the stochastic state space model with additive noise:

$$\mathcal{X}(k+1) = f(\mathcal{X}(k)) + \mathcal{W}(k), \tag{2.5}$$

$$\mathcal{Y}(k) = h(\mathcal{X}(k)) + \mathcal{V}(k), \tag{2.6}$$

where $\mathcal{X}(k) \in \mathbb{R}^n$ is the state vector at time k, $\mathcal{W}(k) \in \mathbb{R}^n$ is the process noise, $\mathcal{V}(k) \in \mathbb{R}^r$ is the observation noise and $\mathcal{Y}(k) \in \mathbb{R}^r$ is the noisy observation of the system. It is conventionally assumed that the distribution of $\mathcal{X}(k)$ is Gaussian. Here, $\mathcal{W}(k)$ and $\mathcal{V}(k)$ are often assumed to be independent. Then, an alternative description of the model in terms of conditional densities $p(\mathcal{X}(k+1)|\mathcal{X}(k))$ and $p(\mathcal{Y}(k)|\mathcal{X}(k))$ is easily derived. A number of approximate filtering techniques given, in Section 2.3.2, rely on (2.5)-(2.6) with a Gaussian assumption regarding $\mathcal{W}(k)$ and $\mathcal{V}(k)$. The related algorithms boil down to approximately computing expected values with respect to Gaussian densities.

2.2.3 The Continuous-discrete Time Linear System

The process equation can be described by the stochastic differential equation

$$d\mathcal{X}(t) = (A\mathcal{X}(t) + B)dt + U_w d\mathcal{W}(t). \tag{2.7}$$

The behaviour of the system is observed through noisy measurements $\mathcal{Y}(t_k)$ which are taken at the discrete time instant $t_k = kT$ (T is the measurement sampling interval)

$$\mathcal{Y}(t_k) = C\mathcal{X}(t_k) + D + U_v\mathcal{V}(t_k), \tag{2.8}$$

where $\mathcal{X}(t)$ is an n-dimensional state of the system at any time t, $\mathcal{Y}(t_k) \in \mathbb{R}^r$ is the measurement at t_k^{th} time instant with A, C, U_w and U_v are given constant matrices and B and D are given constant vectors of compatible dimensions. $\mathcal{W}(t) \in \mathbb{R}^n$ is a standard Wiener process with increment $d\mathcal{W}(t)$ and $\mathcal{V}(k) \in \mathbb{R}^r$ is the measurement noise and zero mean, i.i.d. random vectors with identity covariance matrix $\mathcal{I}$. The noise signals $\mathcal{W}(t)$ and $\mathcal{V}(t_k)$ are uncorrelated with each other. The initial state is a random vector with a known mean and covariance matrix $\mathbb{E}[\mathcal{X}(0)] = \hat{\mathcal{X}}(0)$ and $\mathbb{E}[(\mathcal{X}(0)-\hat{\mathcal{X}}(0))(\mathcal{X}(0)-\hat{\mathcal{X}}(0))^\top] = P(0)$, respectively. $\mathcal{X}(0)$, $\mathcal{W}(t)$ and $\mathcal{V}(t_k)$ are mutually independent.

Dynamics of many continuous time systems can be represented by their discrete time approximation with acceptable accuracy with small enough time steps, as given in chapter 5. Now the problem is to estimate the states of this discrete time process. When the state-space model is known, one of the major contributions made to solving this problem is the Kalman filter, which provides the optimal least-squares estimator in Gaussian linear systems. Another alternative to continuous-discrete filtering is to use an ordinary differential equation (ODE) for the evolution of the moments of the state between the measurement sampling times to yield $\hat{\mathcal{X}}(k+1|k)$ and $P(k+1|k)$ and then use a discrete update equation once $\mathcal{Y}(k+1)$ is measured.

In this chapter, we provide the reader with a number of tools to approach the filtering problem. In Section 2.3.1, the Kalman filter is introduced as an exact and optimal estimator for linear systems with known noise statistics. In Section 2.3.2, we introduce algorithms that build on the Kalman filter but consider nonlinear state space models.

2.3 Gaussian Filtering Techniques

2.3.1 The Kalman Filter

One of the most common approaches to latent state estimation in dynamical systems is the Kalman filter, first published in 1960 [1] for the discrete-time case, which provides an optimal solution in the sense of MV for linear systems. The Kalman filter is a recursive filter which estimates the true states of the dynamics of a system by combining model predictions with noisy observations. His work is extended to deal with the continuous-time case in [3], the filter is called the Kalman Bucy filter in that case. It has numerous application in technology, including guidance, navigation and control of vehicles, particularly aircraft and spacecraft [4]. As an example, in radar-based tracking, one may measure the bearing between a fixed axis and a target and the distance between the target and the radar location, and infer the coordinates and velocity of the target.

Furthermore, the Kalman filter is widely applied in time series analysis used in fields such as signal processing and econometrics. An introduction to the general idea of the Kalman filter can be found in [5] and the references therein. More extensive accounts of the Kalman filter's history are given in [6] and [7]. Anderson and Moore [8] produced an early standard text with a focus on discrete-time systems, and provide a number of alternative derivations and many extensions to the Kalman filter.

In general, the Kalman filter addresses the general problem of trying to estimate the state $\mathcal{X} \in \mathbb{R}^n$ of a discrete time process that is governed by the linear stochastic difference equation

$$\mathcal{X}(k+1) = A\mathcal{X}(k) + B + U_w\mathcal{W}(k); \quad k \geq 0 \tag{2.9}$$

with a measurement $\mathcal{Y} \in \mathbb{R}^r$

$$\mathcal{Y}(k) = C\mathcal{X}(k) + D + U_v\mathcal{V}(k), \tag{2.10}$$

where $\mathcal{W}(k) \in \mathbb{R}^n$ and $\mathcal{V}(k) \in \mathbb{R}^r$. The random variables $\mathcal{W}(k)$ and $\mathcal{V}(k)$ represent the process and measurement noise, respectively. They are independent and are assumed to have normal probability distributions with zero mean and unit variance:

$$p(\mathcal{W}) \sim \mathcal{N}(0, I),$$

$$p(\mathcal{V}) \sim \mathcal{N}(0, I),$$

Further, the noise variables are not correlated forward or backward in time so that

$$\mathbb{E}[\mathcal{V}(k)\mathcal{V}^\top(j)] = 0 \qquad if \qquad k \neq j,$$

$$\mathbb{E}[\mathcal{V}(k)\mathcal{V}^\top(j)] = I \qquad if \qquad k = j,$$

and

$$\mathbb{E}[\mathcal{W}(k)\mathcal{W}^\top(j)] = 0 \qquad if \qquad k \neq j,$$

$$\mathbb{E}[\mathcal{W}(k)\mathcal{W}^\top(j)] = I \qquad if \qquad k = j.$$

We further assume that $\mathcal{V}(k)$ and $\mathcal{W}(k)$ are uncorrelated so that

$$\mathbb{E}[\mathcal{V}(k)\mathcal{W}^\top(j)] = 0.$$

The $n \times n$ matrix A in the difference equation (2.9) relates the state at time step k to the state at step $k+1$, in the absence of either a driving function or process noise. The $r \times n$ matrix C in the measurement equation (2.10) relates the state at time step k to the measurement $\mathcal{Y}(k)$. B, D, U_w and U_v are constant matrices.

The initial state, $\mathcal{X}_0$, is a Gaussian random vector with mean

$$\mathbb{E}[\mathcal{X}_0] = \hat{\mathcal{X}}_0,$$

and the estimation error covariance matrix

$$\mathbb{E}[(\mathcal{X}_0 - \hat{\mathcal{X}}_0)(\mathcal{X}_0 - \hat{\mathcal{X}}_0)^T] = P_{x_0}.$$

The Kalman filter algorithm

Kalman filter equations can be stated in many different forms of which we present a scheme with alternating time and measurement updates. For the system (2.9)-(2.10), assume that the conditional expectation $\hat{\mathcal{X}}(k|k)$ and its covariance matrix $P_{xx}(k|k)$ at time k, which were obtained by processing all measurements $\mathcal{Y}(1:k)$, are known. The Kalman filtering algorithm for finding conditional moments at the next time $(k+1)$ proceeds as follows [9]:

$$\hat{\mathcal{X}}(k+1|k) = A\hat{\mathcal{X}}(k|k) + B, \tag{2.11}$$

$$P_{xx}(k+1|k) = AP_{xx}(k|k)A^\top + U_w U_w^\top, \tag{2.12}$$

$$\hat{\mathcal{V}}(k+1) = \mathcal{Y}(k+1) - C\hat{\mathcal{X}}(k+1|k) - D, \tag{2.13}$$

$$P_{xv}(k+1|k) = AP_{xx}(k+1|k)C^\top, \tag{2.14}$$

$$P_{vv}(k+1|k) = CP_{xx}(k+1|k)C^\top + U_v U_v^\top, \tag{2.15}$$

$$\hat{\mathcal{X}}(k+1|k+1) = \hat{\mathcal{X}}(k+1|k) + P_{xv}(k+1|k)P_{vv}^{-1}(k+1|k)\hat{\mathcal{V}}(k+1), \tag{2.16}$$

$$P_{xx}(k+1|k+1) = P_{xx}(k+1|k) - P_{xv}(k+1|k)P_{vv}^{-1}(k+1|k)P_{xv}^\top(k+1|k), \tag{2.17}$$

where $\hat{\mathcal{X}}(k+1|k)$ denotes the optimal estimate of $\mathcal{X}$ at time $k+1$, given the measurements and other available values up to time k. The calculation of $\hat{\mathcal{X}}(k+1|k)$ can be interpreted as performing the expectation of (2.9) over the random variables $\mathcal{X}(k)$ and $\mathcal{W}(k)$. $\bar{P}(k+1|k)$ represents the related covariance which is, again, an expected value. $\hat{\mathcal{V}}(k+1)$ in (2.13) is called innovation, which is the discrepancy between the observed output and its prediction. $P_{vv}(k+1|k)$ represents the covariance matrix of innovation. $P^\top$ denotes the transpose of matrix P. To initialise the procedure, it is assumed that $\hat{\mathcal{X}}_0$ and P_{x_0} are known, and proceeding with a time update.

Equation (2.16) is an optimal linear filter, in the sense that it yields the MV over all linear filters, even when $\mathcal{V}(k)$ and $\mathcal{W}(k)$ are not Gaussian. When $\mathcal{V}(k)$ and $\mathcal{W}(k)$ are Gaussian, $\hat{\mathcal{X}}(k+1|k)$ is the conditional mean estimator for $\mathcal{X}(k+1)$, given $\mathcal{Y}(k)$. In fact, equation (2.16) may be derived using a standard conditional mean relationship for two Gaussian variables $\mathcal{X}$ and $\mathcal{Y}$:

$$\mathbb{E}(\mathcal{X}|\mathcal{Y}) = \mathbb{E}(\mathcal{X}) + P_{XY}P_{YY}^{-1}(\mathcal{Y} - \mathbb{E}(\mathcal{Y})),$$

where P_{YY} and P_{XY} are covariance matrices.

2.3.2 Nonlinear Filtering Problem

In most applications of interest, the state transition and the observation equations are nonlinear and the Kalman filter cannot be directly applied. As a result, a wide variety of approximate nonlinear filters have been considered, such as the the EKF as discussed in [10], the UKF proposed in [11], EnKF proposed in [12] and the QKF as reviewed in [13]. In the EKF, the system state distribution and all relevant noise densities are approximated by Gaussian random variables (GRVs), which are then propagated analytically through a first-order linearisation of the nonlinear system. On the other hand, in the UKF the state distribution is again approximated by a GRV as with a EKF, but is now represented using a minimal set of carefully chosen weighted sample points. The QKF uses numerical quadrature for finding the mean and variance of the latent state. The most common approach is to use the EKF, which simply linearizes the nonlinear equations around the current estimate so that the traditional linear Kalman filter can be applied.

2.3.3 The Extended Kalman Filter

In the EKF [14], the PDF is propagated through a linear approximation of the system around the operating point at each time instant. In doing so, the EKF requires Jacobian matrices. The EKF can be improved by using the current estimate of the state vector to linearize the measurement equation in an iterative model. This approach is known as the Iterated EKF (IEKF) [15].

The EKF Algorithm

Let us assume that our process again has a state vector $\mathcal{X} \in \mathbb{R}^n$, but that the process is now governed by the nonlinear stochastic difference equation (2.5) with a measurement (2.6).

The fundamental concept of the EKF involves the notion that the true state is sufficiently close to the estimated state. Therefore, it provides approximation of the mean and covariance which are accurate up to at least the first term of their Taylor series expansions. For a discussion of and more detail on the EKF, see [2].

Consider the Taylor series expansion of the nonlinear functions f and h around the estimates $\hat{\mathcal{X}}(k|k-1)$ of the states $\mathcal{X}(k)$.

$$\begin{aligned} f(\mathcal{X}(k)) &\approx f(\hat{\mathcal{X}}(k|k-1)) + \frac{\partial f(\hat{\mathcal{X}}(k|k-1))}{\partial \mathcal{X}}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k-1)) + \dots \\ h(\mathcal{X}(k)) &\approx h(\hat{\mathcal{X}}(k|k-1))) + \frac{\partial h(\hat{\mathcal{X}}(k|k-1)),)}{\partial \mathcal{X}}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k-1)) + \dots \end{aligned} \tag{2.18}$$

Equation (2.18) gives a linear approximation of the original nonlinear state space system. Using only the linear expansion terms, it is easy to derive the update equations for the mean $\hat{\mathcal{X}}$ and covariance P of the Gaussian approximation to the posterior distribution of the states. The EKF algorithm for the system (2.5)-(2.6) is presented below [14]:

- Initialization at $k = 0$:

$$\begin{aligned} \hat{\mathcal{X}}_0 &= \mathbb{E}[\mathcal{X}_0], \\ P_{x_0} &= \mathbb{E}[(\mathcal{X}_0 - \hat{\mathcal{X}}_0)(\mathcal{X}_0 - \hat{\mathcal{X}}_0)^\top], \\ P_w &= \mathbb{E}[(\mathcal{W} - \bar{\mathcal{W}})((\mathcal{W} - \bar{\mathcal{W}}))^T], \\ P_v &= \mathbb{E}[(\mathcal{V} - \bar{\mathcal{V}})(\mathcal{V} - \bar{\mathcal{V}})^T], \end{aligned}$$

- For $k \geq 1$:

(1) Prediction step.

(a) Compute the process model Jacobians:

$$F_x(k) = \nabla_x f(\mathcal{X}, \bar{\mathcal{W}})|_{\mathcal{X}=\hat{\mathcal{X}}(k|k-1)},$$

$$G_w = \nabla_w f(\hat{\mathcal{X}}(k|k-1), \mathcal{W})|_{\mathcal{W}=\bar{\mathcal{W}}}.$$

(b) Compute predicted state mean and covariance (time update)

$$\hat{\mathcal{X}}^-(k) = f(\hat{\mathcal{X}}(k|k-1), \bar{\mathcal{W}}),$$

$$P_x^-(k) = F_x(k)P_x(k)F_x^T(k) + G_w P_w {G_w}^T.$$

(2) Correction step.

(a) Compute observation model Jacobians:

$$H_x(k) = \nabla_x h(\mathcal{X}, \bar{\mathcal{V}})|_{\mathcal{X}=\hat{\mathcal{X}}^-(k)},$$

$$D_v = \nabla_v h(\hat{\mathcal{X}}^-(k), \mathcal{V})|_{\mathcal{V}=\bar{\mathcal{V}}}.$$

(b) Update estimates with latest observation(measurement update)

$$K_k = P_x^-(k){H_x}^T(k)(H_x(k)P_x^-(k){H_x}^T(k) + D_v P_n {D_v}^T)^{-1},$$

$$\hat{\mathcal{X}}(k) = \hat{\mathcal{X}}^-(k) + K_k[\mathcal{Y}(k) - h(\hat{\mathcal{X}}^-(k), \bar{\mathcal{V}})],$$

$$P_x(k) = (I - K_k H_x(k))P_x^-(k).$$

More detail on this algorithm can be found in [14].

Unfortunately, EKF often cannot be applied in practical applications of nonlinear estimation techniques. It has been successfully applied in [16] and [17], but fails in others [11] for a variety of reasons, including the existence of Jacobians, their computational complexity and stability of the resulting system. Therefore, the EKF has two important potential drawbacks [18]. First, the derivation of the Jacobian matrices, the linear approximators to the nonlinear functions, can cause implementation difficulties. Second, these linearisations can lead to filter instability if the time step intervals are not sufficiently small. However, in the EKF, the system state distribution and all relevant noise densities are approximated by GRVs, which are then propagated analytically through a first-order linearisation of the nonlinear system. This can introduce large errors in the true posterior mean and covariance of the transformed GRV, which may lead to sub-optimal performance and ,sometimes, divergence of the filter.

To overcome this limitation, Julier, Uhlmann and Durrant-Whyte introduced a new filtering algorithm called the UKF [19].

2.3.4 The Unscented Kalman Filter

The UKF is a superior alternative to the EKF for a variety of estimation and control problems. The UKF addresses the EKF's drawbacks by using a deterministic sampling approach. The state distribution is again approximated by a GRV as an EKF, but is now represented using a minimal set of carefully chosen weighted sample points. These sample points completely capture the true mean and covariance of the GRV, and when propagated through the true nonlinear system, capture the posterior mean and covariance accurately to the second order (Taylor series expansion) for any nonlinearity [20]. Furthermore, the UKF has substantial advantages over the EKF both in implementation and performance and requires no analytic differentiation or Jacobians, as which are necessary when using the EKF.

The UKF algorithm has found a number of applications in many fields such as communication [11], engineering [21, 19] and finance [22]. The superior performance of the UKF as compared to that of the EKF has been demonstrated in many applications [18, 23] and [14]. Previous research has found that the UKF leads to more accurate results than the EKF and that in particular it generates much better estimates of the covariance of the state. The reason for this greater accuracy is that the UKF can predict the state estimate and error covariance to the fourth order accuracy while the EKF only predicts up to the second order for the state estimate and the fourth order for the error covariance [19].

The UKF is based on a nonlinear transformation which nonlinearly propagates the mean and covariance information and is known as the unscented transformation [11].

The Unscented Transformation

The unscented transformation (UT) is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation. It is based on the intuition that "it is easier to approximate a probability distribution than a nonlinear function" [24].

Consider propagating a random variable $\mathcal{X} \in \mathbb{R}^L$ through an arbitrary nonlinear function $\mathcal{Y} = g(\mathcal{X})$. Assume $\mathcal{X}$ has mean $\hat{\mathcal{X}}$ and covariance P_x. To calculate the first two moments of $\mathcal{Y}$ using UT, we form a set of $2L+1$ sample points $\mathcal{X}_i; i = 0, ..., 2L$ where $\mathcal{X}_i \in \mathbb{R}^L$, with each point being associated with a weight w_i. These sample points are called sigma points. It is important to note that these points are not chosen at random but rather according to some deterministic algorithm. The sigma points are calculated using the following general selection scheme [25]:

$$\mathcal{X}_0 = \hat{\mathcal{X}}, \tag{2.19}$$

$$\mathcal{X}_i = \hat{\mathcal{X}} + \xi(\sqrt{P_x})_i, \quad i = 1, ..., L, \tag{2.20}$$

$$\mathcal{X}_i = \hat{\mathcal{X}} - \xi(\sqrt{P_x})_i, \quad i = L+1, ..., 2L, \tag{2.21}$$

where $\xi = \sqrt{L+\lambda}$ is a scalar scaling factor that determines the spread of the sigma points around $\hat{\mathcal{X}}$, $\lambda = \alpha^2(L+\kappa) - L$ is a compound scaling parameter, L is the dimension of the augmented state vector and $0 < \alpha \leq 1$ is the primary scaling factor determining the extent of the spread of the sigma points around the prior mean. The typical range for α is $1e-3 < \alpha \leq 1$. κ is a tuning parameter and we must choose $\kappa \geq 0$ to guarantee the semi-positive definiteness the covariance matrix; a good default choice is $\kappa = 0$. $(\sqrt{P})_i$ indicates the i^{th} column of the matrix square root of the covariance matrix P.

Once the sigma points are calculated from the prior statistics as shown above, they are propagated through the nonlinear function

$$\mathcal{Y}_i = g(\mathcal{X}_i), \quad i = 0, ..., 2L, \tag{2.22}$$

and the mean and covariance of $\mathcal{Y}$ are approximated using a weighted sample mean and covariance of the posterior sigma points

$$\begin{aligned} \hat{\mathcal{Y}} &\approx \sum_{i=0}^{2L} w_i{}^m \mathcal{Y}_i, \quad \sum w_i{}^m = 1, \\ P_y &\approx \sum_{i=0}^{2L}\sum_{j=0}^{2L} w_{ij}^c \mathcal{Y}_i \mathcal{Y}_j^T, \\ P_{xy} &\approx \sum_{i=0}^{2L}\sum_{j=0}^{2L} w_{ij}^c \mathcal{X}_i \mathcal{Y}_j^T, \end{aligned} \tag{2.23}$$

$$\begin{aligned} w_0^m &= \lambda/(L+\lambda), \\ w_0^c &= w_0^m + (1-\alpha^2+\beta), \\ w_i^c &= w_i^m = 1/[2(L+\lambda)] \quad for \quad i = 1, ..., 2L, \end{aligned} \tag{2.24}$$

where $w_i{}^m$ and $w_i{}^c$ are scalar weights of the mean and covariance calculation associated with the i^{th} point and β is a secondary scaling factor used to emphasize the weighting on the zeroth sigma point for the posterior covariance calculation. β can be used to minimize certain higher-order error terms based on known moments of the prior random variable. For a Gaussian prior, the optimal choice is $\beta = 2$ according to [25].

The UKF Algorithm

Let the system be represented by (2.5) and (2.6); in a UKF the state random variable is redefined as the concatenation of the original state and noise variables: $\mathcal{X}^a(k) = [\mathcal{X}^T(k) \quad \mathcal{W}^T(k) \quad \mathcal{V}^T(k)]^\top$. The sigma points selection scheme (system (2.19)-(2.21)) is applied to this new augmented state random variable to calculate the corresponding sigma points set, $\mathcal{X}_i^a(k)$ where $\mathcal{X}_i^a(k) \in \mathbb{R}^{L_x+L_v+L_n}$.

The UKF algorithm is given below[25]:

- Initialization at $k = 0$:

$$\hat{\mathcal{X}}_0 = \mathbb{E}[\mathcal{X}_0],$$
$$P_{x_0} = \mathbb{E}[(\mathcal{X}_0 - \hat{\mathcal{X}}_0)(\mathcal{X}_0 - \hat{\mathcal{X}}_0)^T],$$
$$\hat{\mathcal{X}}_0^a = \mathbb{E}[\mathcal{X}_0^a] = [\hat{\mathcal{X}}_0^T \quad \bar{\mathcal{W}}_0^T \quad \bar{\mathcal{V}}_0^T],$$
$$P_0^a = \mathbb{E}[(\mathcal{X}_0^a - \hat{\mathcal{X}}_0^a)(\mathcal{X}_0^a - \hat{\mathcal{X}}_0^a)^T = \begin{bmatrix} P_{x_0} & 0 & 0 \\ 0 & R_w & 0 \\ 0 & 0 & R_v \end{bmatrix}.$$

- For $k \geq 1$:

1. Calculate sigma-points:

$$\tilde{\mathcal{X}}^a(k-1) = [\hat{\mathcal{X}}^a(k|k-1) \quad \hat{\mathcal{X}}^a(k|k-1)+\xi(\sqrt{P^a(k|k-1)}) \quad \hat{\mathcal{X}}^a(k|k-1)-\xi(\sqrt{P^a(k|k-1)})].$$

2. Time-update equations:

$$\mathcal{X}^x(k|k-1) = f(\mathcal{X}^x(k-1), \mathcal{X}^w(k-1)),$$
$$\hat{\mathcal{X}}^-(k) = \sum_{i=0}^{2L} w_i^m \mathcal{X}_i^x(k|k-1),$$

$$P^-_{x(k)} = \sum_{i=0}^{2L} w_i^c(\mathcal{X}_i^x(k|k-1) - \hat{\mathcal{X}}^-(k))(\mathcal{X}_i^x(k|k-1) - \hat{\mathcal{X}}^-(k))^T.$$

3. Measurement-update equations:

$$\mathcal{Y}(k|k-1) = h(\mathcal{X}^x(k|k-1), \mathcal{X}^v(k-1)),$$

$$\hat{\mathcal{Y}}^-(k) = \sum_{i=0}^{2L} w_i^m \mathcal{Y}_i(k|k-1),$$

$$P_{\tilde{y}(k)} = \sum_{i=0}^{2l} w_i^c(\mathcal{Y}_i^-(k|k-1) - \hat{\mathcal{Y}}^-(k))(\mathcal{Y}_i(k|k-1) - \hat{\mathcal{Y}}^-(k))^T,$$

$$P_{x(k)y(k)} = \sum_{i=0}^{2L} w_i^c(\mathcal{X}_i^x(k|k-1) - \hat{\mathcal{X}}^-(k))(\mathcal{Y}_i(k|k-1) - \hat{\mathcal{Y}}^-(k))^T,$$

$$K_k = P_{x(k)y(k)} P^{-1}_{\tilde{y}(k)},$$

$$\hat{\mathcal{X}}(k) = \hat{\mathcal{X}}^-(k) + K_k(\mathcal{Y}(k) - \hat{\mathcal{Y}}^-(k)),$$

$$P_{x(k)} = P^-_{x(k)} - K_k P_{\tilde{y}(k)} K_k^T,$$

where $\tilde{\mathcal{X}} = [(\mathcal{X}^x)^\top \quad (\mathcal{X}^w)^\top \quad (\mathcal{X}^v)^\top]^\top$, w_i^c and w_i^m are as defined in (2.24). Many extensions, generalizations and developments followed Julier and Uhlmann's basic work [19]. They developed the UKF algorithm by extending the number of sigma points in order to capture the first four moments of a Gaussian distribution in [21]. This algorithm was improved by Wan and Van der Merwe [20], who extended the use of the UKF to a broader class of nonlinear estimation problems, including nonlinear system identification, training of neural networks and dual estimation problems. Julier, Uhlmann and Durrant-Whyte in [24] described a new approach for generalizing the Kalman filter to nonlinear systems. They approximated the first three moments of the prior distribution accurately using a set of samples. The algorithm predicts the mean and covariance accurately up to the third order. Ponomareva, Date and Wang [26] also introduced a new UKF for nonlinear multivariate time series. This UKF generates sample points and corresponding probability weights which are modified at each step. These sample points and weights match exactly the predicted values of average marginal skewness and average marginal kurtosis of the unobserved state variables, as well as their means and the covariance matrix.

In order to reduce computational errors, a square root formulation of the UKF which propagates the mean and square root of the covariance matrix rather than the covariance matrix itself has been developed [27]. Hermoso-Carazo and Linares-Perezad [28, 29] addressed the least squares filtering problem for nonlinear systems with uncertain observations. Reformulations of the UKF algorithm have been proposed in [30] by adding constraints using the traditional UKF approach. Gustafsson and Hendeby, who pointed out the connection between the EKF and the UKF, showed that the sigma point function evaluation can be used in the classical EKF in place of an explicitly linearized model [31].

In the literature, various modifications to both the EKF and the UKF have led to improved accuracy or reduced computational complexity have been proposed in [27] and [9], among others. All the approximate filtering methods have their own advantages and disadvantages and provide different levels of trade-offs in terms of computational complexity and estimation accuracy. [32] and [33] provide reviews of filtering applications.

The methods described above, such as the EKF and the UKF, do not have provable optimality properties for non-Gaussian disturbances and nonlinear systems. In contrast, a Bayesian filtering methodology which depends on sampling from appropriate candidate distributions and adjusting the probability weights to suit measurements can recover the true posterior distribution $\mathbb{P}(\mathcal{X}(k)|\mathcal{Y}(k))$, as the number of samples tend to infinity. This kind of Bayesian recursive filter is called a *particle filter* (PF). It is discussed widely in the literature; see [34] for example. With a PF, a discrete approximation of the required density functions is obtained using a set of samples from an appropriate PDF. These density functions and the corresponding probability weights are used to compute the conditional moment estimates. Given their practical and theoretical importance, PFs are discussed in more detail in the next section, along with a numerical illustration.

2.4 Particle Filters

It is well known that when systems are non-Gaussian and/or nonlinear, there are not many methods available to reach the filtering goal (i.e., estimating the distribution of the true state of a process). To handle these problems, PFs, introduced in [34], have become a very popular class of numerical methods for the solution of optimal estimation problems in nonlinear, non-Gaussian models. Thus, the motivation for using PFs lies in their ability to estimate sequentially the densities of unknowns of non-Gaussian and/or nonlinear systems based on the concept of sequential importance sampling and the use of Bayesian theory. It is a technique for implementing a recursive Bayesian filter by Monte Carlo simulations. The key idea of a PF is to try to represent the required PDF by a set of random samples with associated weights that can be interpreted as probability masses, and to compute estimates based on these samples and weights. The set of samples are called *particles* and its location varies with time in a random way. The particles and weights form a discrete random measure.

Many versions of PF algorithms have appeared independently in several fields under such names as condensation [35], sequential Monte Carlo [36], survival of the fittest [37], sequential importance sampling (SIS) with resampling (SIR) [38], [39], bootstrap filters [40], *etc.* The literature features several reviews of PF. For instance, in [41] a short introduction to PF is given and many problems in wireless communications are discussed. In [42] Van and Peter present an excellent overview of PF theory, including several examples from the geoscience. The fundamentals of particle filtering and its most important implementations are discussed in [43]. Practical applications include target tracking [44], estimation of stochastic volatility [45], geosciences [46] and blind deconvolution of digital communications channels [47]. In [48] a few convergence results on particle filtering methods are reviewed. Numerous statistical improvements for PFs have been proposed, and some important theoretical properties have been established. For example, in order to reduce the computational cost of implementing PFs, an improved PF is proposed in [49]. In [50] a recursive on-line algorithm based on rejection sampling is provided and improved versions suggested. A fully nonlinear PF that can be applied to higher dimensional problems appears in [51]. The method is applied for the highly nonlinear three-dimensional Lorenz model [52] with only partial observations of the state vector with only three particles and the much more complex 40-dimensional Lorenz model [53] using 20 particles. The same problem has been studied before [54] using standard a PF with resampling and tens of thousands of particles were needed. In general, the underlying principle of particle filtering is that the relevant distributions (both prior and posterior) are approximated by discrete random measures composed of particles (samples from the space of the unobserved variables) and weights assigned to those particles.

2.4.1 Basic particle filters

In many applications we are interested in estimating recursively in time the unobserved state $\mathcal{X}(k)$ from the observations $\mathcal{Y}(k)$, where the state model is (2.5) and the observation model is (2.6). As already pointed out, the central idea of PF is about approximating $\mathcal{X}(k)$ given the measurement $\mathcal{Y}(k)$, which we will represent as $\mathcal{X}(k|k)$, with the use of discrete random measure $\chi^{(m)}(k) = \{\mathcal{X}^{(m)}(k), W^{(m)}(k)\}_{m=1}^{M}$, where $\mathcal{X}^{(m)}$ are the particles, $W^{(m)}$ are the probability weights and M is the number of particles. The algorithm starts by drawing an initial independent sample $\mathcal{X}_0^{(1)}, \ldots, \mathcal{X}_0^{(M)}$; all particles have equal weight, which are then inserted into the system equation (2.5) - (2.6). The set must be updated by the new measurement $\mathcal{Y}(1)$. Suppose that at time $k-1$, we know the observations up to $k-1$ and the a posteriori PDF $p(\mathcal{X}(k-1)|\mathcal{Y}(k-1))$. Once $\mathcal{Y}(k)$ becomes available, we would like to update $p(\mathcal{X}(k-1)|\mathcal{Y}(k-1))$ and modify it to $p(\mathcal{X}(k)|\mathcal{Y}(k))$. To achieve this, we formally write

$$p(\mathcal{X}(k)|\mathcal{Y}(k)) \propto p(\mathcal{Y}(k)|\mathcal{X}(k))p(\mathcal{X}(k)|\mathcal{Y}(k-1)). \tag{2.25}$$

The first factor on the right of the proportionality sign is the likelihood function of the unknown state, and the second factor is the predictive density of the state. In order to estimate the posterior distribution $p(\mathcal{X}(k)|\mathcal{Y}(k)$, the particles drawn from it are used. If we sample a large number of particles M from $p(\mathcal{X}(k)|\mathcal{Y}(k))$, we will be able to estimate $E(h(\mathcal{X}(k))$ with arbitrary accuracy. In practice, however, the problem is that we often cannot draw samples directly from the a posteriori PDF, i.e., $p(\mathcal{X}(k)|\mathcal{Y}(k))$. An attractive alternative is to use the concept of importance sampling [58], which is based on another function for drawing particles. This function is called the importance sampling function or proposal distribution, and we denote it by $q(\mathcal{X}(k)|\mathcal{Y}(k))$. This importance function minimizes the variance of the weights. Given the particles, which are obtained by sampling from the proposal distribution $q(\mathcal{X}(k)|\mathcal{X}(k-1), \mathcal{Y}(k))$, a weighted approximation to the posterior density at k is given by

$$p(\mathcal{X}(k)|\mathcal{Y}(k) \approx \sum_{m=1}^{M} W^{(m)}(k)\delta(\mathcal{X}(k) - \mathcal{X}^{(m)}(k)), \tag{2.26}$$

where

$$W^{(m)}(k) \propto \frac{p(\mathcal{X}^{(m)}(k)|\mathcal{Y}(k))}{q(\mathcal{X}^{(m)}(k)|\mathcal{Y}(k))}, \tag{2.27}$$

and $\delta(\cdot)$ is the Dirac delta function. From the law of large numbers $\sum_{m=1}^{M} W^m(k)h(\mathcal{X}^m(k)) \rightarrow \mathbb{E}(h(\mathcal{X}(k)))$ as $M \rightarrow \infty$. If the importance density is chosen to factorise such as

$$q(\mathcal{X}(k+1)|\mathcal{Y}(k+1)) = q(\mathcal{X}(k+1)|\mathcal{X}(k),\mathcal{Y}(k+1))q(\mathcal{X}(k)|\mathcal{Y}(k)), \quad (2.28)$$

we can augment the trajectory $\mathcal{X}^{(m)}(k)$ with $\mathcal{X}^{(m)}(k+1)$, where

$$\mathcal{X}^{(m)}(k+1) \sim q(\mathcal{X}(k+1)|\mathcal{X}^{(m)}(k),\mathcal{Y}(k+1)).$$

To derive the weight update equation, $p(\mathcal{X}(k+1)|\mathcal{Y}(k+1))$ is first expressed in terms of $p(\mathcal{X}(k)|\mathcal{Y}(k))$, $p(\mathcal{Y}(k+1)|\mathcal{X}(k+1))$ and $p(\mathcal{X}(k+1)|\mathcal{X}(k))$:

$$p(\mathcal{X}(k+1)|\mathcal{Y}(k+1)) \propto p(\mathcal{Y}(k+1)|\mathcal{X}(k+1))p(\mathcal{X}(k+1)|\mathcal{X}(k))p(\mathcal{X}(k)|\mathcal{Y}(k)). \quad (2.29)$$

By substituting (2.28) and (2.29) into (2.27), the weight update equation is then shown to be

$$W^{(m)}(k+1) = W^{(m)}(k)\frac{p(\mathcal{Y}(k+1)|\mathcal{X}^{(m)}(k+1))p(\mathcal{X}^{(m)}(k+1)|\mathcal{X}^{(m)}(k))}{q(\mathcal{X}^{(m)}(k+1)|\mathcal{X}^{(m)}(k),\mathcal{Y}(k+1))}, \quad (2.30)$$

where $p(\mathcal{Y}(k+1)|\mathcal{X}^{(m)}(k+1))$ represents the probability density of the observations given the model state $\mathcal{X}^m$, which is determined by the probability density of the observation noise and is often taken as a Gaussian, and $p(\mathcal{X}^{(m)}(k+1)|\mathcal{X}^{(m)}(k))$ represents the state transition density. $q(\mathcal{X}^{(m)}(k+1)|\mathcal{X}^{(m)}(k),\mathcal{Y}(k+1))$ represents the proposal distribution and $p(\mathcal{X}(k-1)|\mathcal{X}(k))$ is a common choice for the proposal distribution.

A major problem with particle filtering is that the discrete random measure degenerates quickly. In other words, all the particles except one have negligible weights. This means that the statistical information in the ensemble is lost; effectively only one particle has all the information available to us. The degeneracy implies that the performance of the PF will deteriorate. Degeneracy, however, can be reduced by using good importance sampling functions and resampling [43]. The basic idea of resampling is to eliminate particles with small weights and replicate particles with large weights, such that we end up with an ensemble of particles with equal weight again. A suitable measure of degeneracy of the algorithm is the effective sample size M_{eff} introduced in [39] and defined as

$$M_{\text{eff}} = \frac{M}{1 + Var(W^{\star m}(k))}$$

where $W^{\star m}(k) = p(\mathcal{X}^m(k)|\mathcal{Y}(k))/q(\mathcal{X}^{(m)}(k)|\mathcal{X}^{(m)}(k-1), \mathcal{Y}(k))$ is referred to as the "true weight". One cannot evaluate M_{eff} exactly but, an estimate $\hat{M}_{\text{eff}}$ of M_{eff} is given by

$$\hat{M}_{\text{eff}} = \frac{1}{\sum_{m=1}^{M}(W^m(k))^2},$$

with $W^m(k)$ being the normalised weight corresponding to the m-th particle at time instant k. When $\hat{M}_{\text{eff}}$ is below a fixed threshold, resampling is carried out. Several ways to perform the resampling exist; see for example [39]. The most widely used class of resampling techniques involves resampling that is random and has a uniform distribution. That is all weights are put on the interval $[0;1]$, and a random number from the uniform density over $[0;1/M]$ is chosen. That number is laid onto the unit interval, and the weight it points to is the first resampled particle. Then, $1/M$ is added to the random number, and the weight to which it points denotes the second resampled particle. This process is repeated to generate N resampled particles, all with equal weight. From there we start the model integrations forward in time again.

In general, particle filtering algorithm can be summarized as follows [34]. Suppose that at time step k a discrete probability measure $\{\mathcal{X}^{(m)}(k), W^{(m)}(k)\}$, $m = 1, 2, ..., M$, is specified and a proposal density $q(\mathcal{X}(k+1)|\mathcal{X}(k), \mathcal{Y}(k+1))$ is given. At time steps $(k+1)$, the following sequence of operations gives us the probability measure $\{(\mathcal{X}^{(m)}(k+1), W^m(k+1))\}$:

- At time $k+1$, measure $\mathcal{Y}(k+1)$;

- Sample particles $\mathcal{X}^m(k+1)$ from proposal density $q(\mathcal{X}(k+1|\mathcal{X}(k), \mathcal{Y}(k+1));$

- Compute the importance weights $W^{(m)}(k+1)$ according to (2.30);

- Normalize the importance weights as the following

$$\widetilde{W}^{(m)}(k+1) = \frac{W^{(m)}(k+1)}{\sum_{m=1}^{M} W^{(m)}(k+1)}$$

  ;

- Resample if necessary whenever degeneracy is observed to obtain m equally weighted particles, $\{\bar{\mathcal{X}}^{(m)}, \frac{1}{M}\}_{m=1}^{M}$.

2.4.2 Numerical experiment

In this section, the problem of manoeuvering target tracking [55] with constant but unknown turn rate has been formulated and solved with particle filtering.

Process model

To formulate the problem, we assume an object is manoeuvering with a constant turn rate in a plane parallel to the ground i.e., during manoeuver the hight of the vehicle remains constant. If the turn rate is a known constant, the process model remains linear. However, the constant and known turn rate which needs to be estimated forces the process model to a set of nonlinear equations. The equation of motion of an object in a plane $(X; Y)$ following a coordinated turn model can be described with

$$\ddot{\mathcal{X}} = -\Omega\dot{\mathcal{Y}},$$

$$\ddot{\mathcal{Y}} = \Omega\dot{\mathcal{X}},$$

$$\dot{\Omega} = 0,$$

where $\mathcal{X}$ and $\mathcal{Y}$ represent respectively the positions in the X and Y directions. Ω is the angular rate, which is a constant. The state space representation of the above equations is

$$\dot{\mathcal{X}} = A\mathcal{X} + \mathcal{W},$$

where $\mathcal{X}$ is a state vector defined as $\mathcal{X} = [\mathcal{X} \quad \dot{\mathcal{X}} \quad \mathcal{Y} \quad \dot{\mathcal{Y}}]^\top$. The process noise is added to incorporate the uncertainties in the process equation that arise due to wind speed, variation in turn rate, change in velocity, *etc.* The target dynamics are discretised to obtain the discrete process equation

$$\mathcal{X}(k+1) = F\mathcal{X}(k) + \mathcal{W}(k),$$

where

$$F = \begin{bmatrix} 1 & \frac{\sin(\Omega T)}{\Omega} & 0 & \frac{1-\cos(\Omega T)}{\Omega} \\ 0 & \cos(\Omega T) & 0 & -\sin(\Omega T) \\ 0 & \frac{1-\cos(\Omega T)}{\Omega} & 1 & \frac{\sin(\Omega T)}{\Omega} \\ 0 & \sin(\Omega T) & 0 & \cos(\Omega T) \end{bmatrix}$$

Measurement model

In general, the nonlinear measurement equation can be written as

$$\mathcal{Z}(k) = \gamma(\mathcal{X}(k)) + \mathcal{V}(k).$$

In this problem, we assume both the range and the bearing angle are available from measurements, so the nonlinear function $\gamma(.)$ becomes

$$\gamma(\mathcal{X}(k)) = \begin{bmatrix} \sqrt{\mathcal{X}^2(k) + \mathcal{Y}^2(k)} \\ atan2(\mathcal{Y}(k), \mathcal{X}(k)) \end{bmatrix} + \mathcal{V}(k),$$

where $atan2$ is the four quadrant inverse tangent function. Both $\mathcal{W}(k)$ and $\mathcal{V}(k)$ are white Gaussian noises of zero mean and Q and R covariances respectively and T is sampling time. The process noise covariance Q, is given by

$$Q = q \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} & 0 & 0 \\ \frac{T^2}{2} & T & 0 & 0 \\ 0 & 0 & \frac{T^3}{3} & \frac{T^2}{2} \\ 0 & 0 & \frac{T^2}{2} & T \end{bmatrix},$$

where $T = 0.5$ seconds, $\Omega = -3°/s$ and q is some constant given as $q = 0.1$. $R = diag([\sigma_r^2 \sigma_t^2])$, where $\sigma_r = 120m$ and $\sigma_t = \sqrt{70}mrad$.

- Initialization

The truth state is initialized with $\mathcal{X}(0) = [1000m \quad 30m/s \quad 1000m \quad 0m/s]^\top$. The initial estimate $\hat{\mathcal{X}}(0)$ is generated from a Gaussian distribution of mean $\mathcal{X}(0)$ and covariance $P(0) = diag([200m^2 \quad 20m^2/s^2 \quad 200m^2 \quad 20m^2/s^2])$, respectively.

Figure 2.1: Comparison between true state and particle filters estimate for the first state.

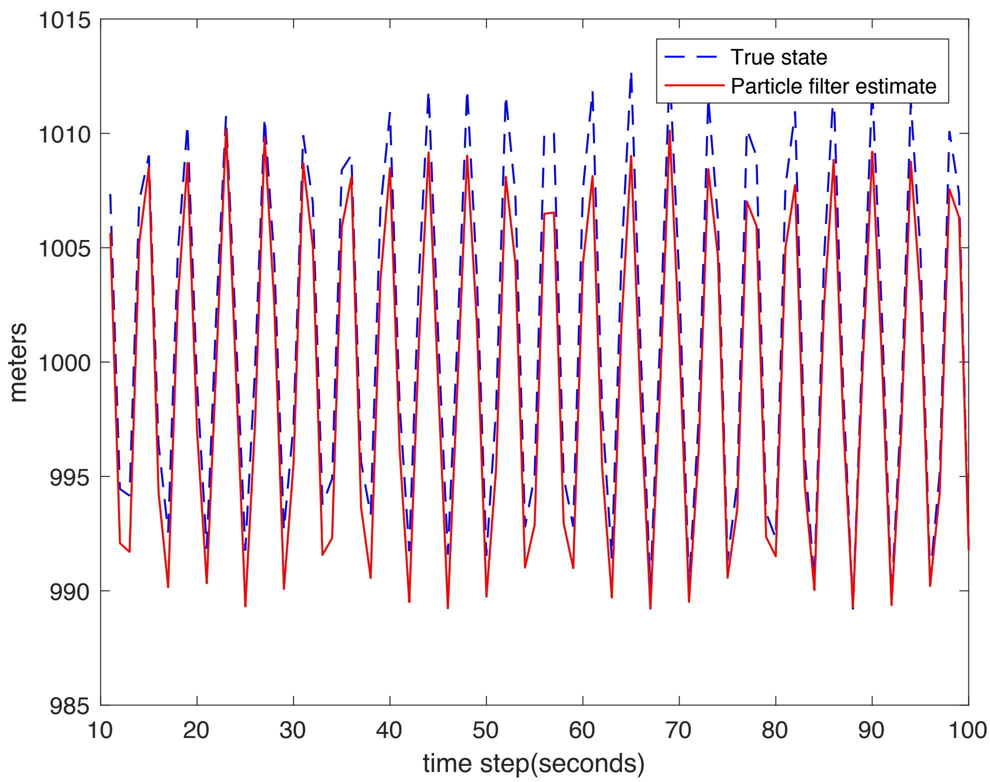
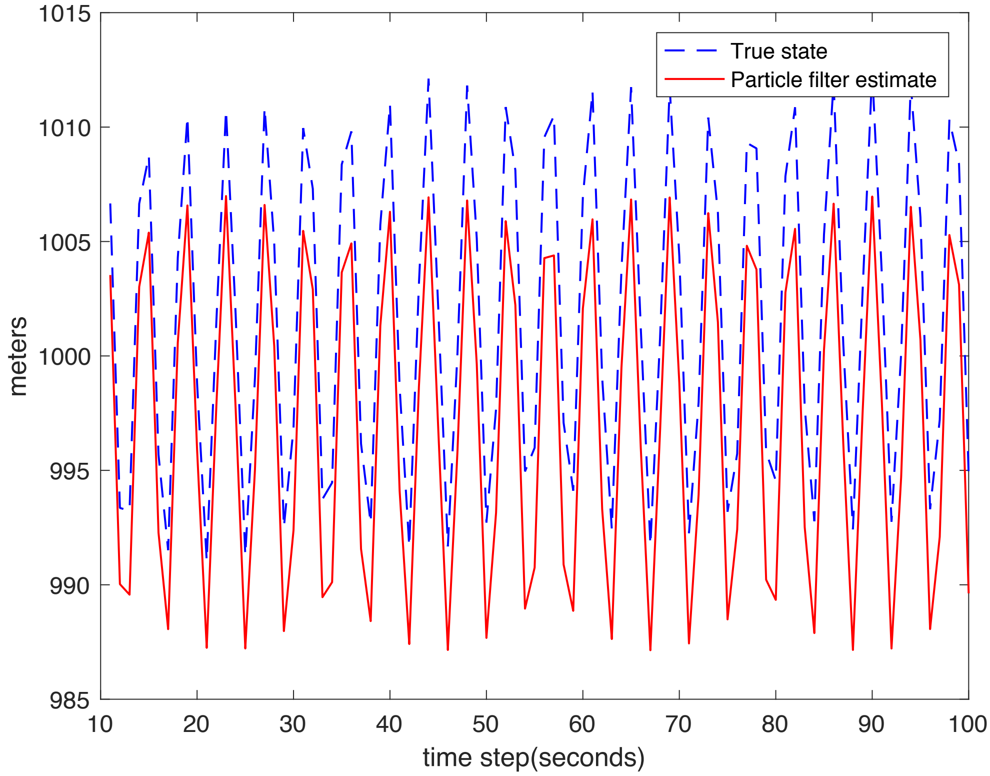


Figure 2.2: Comparison between true state and particle filters estimate for the second state.

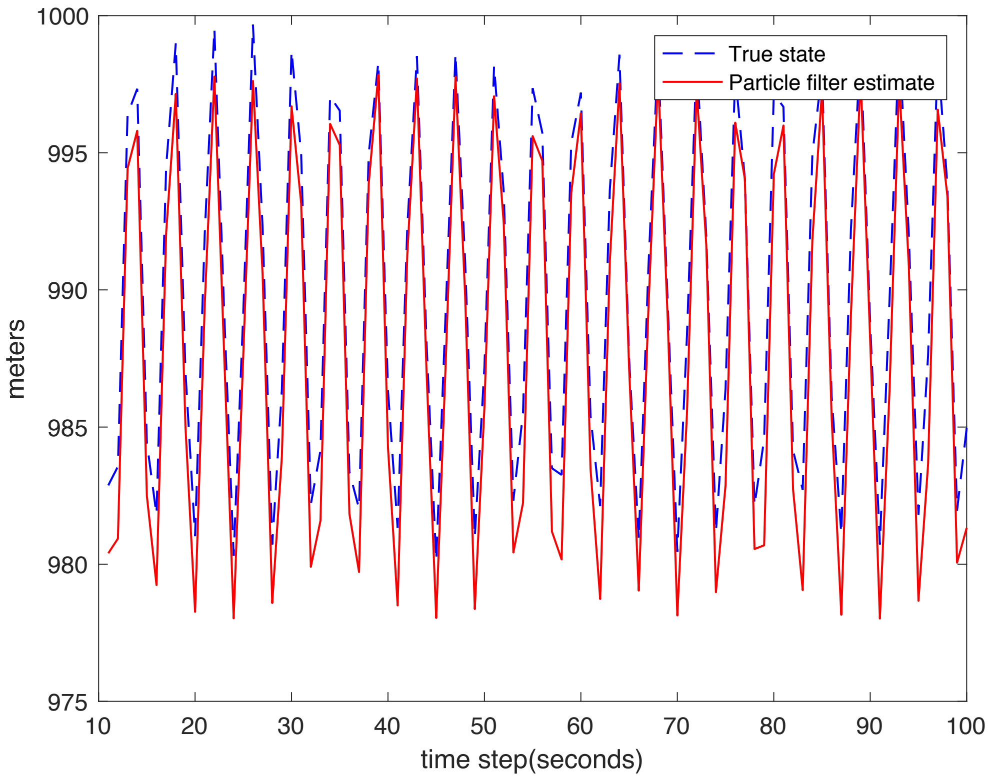


We carried out 50 independent simulations. The steps in s^{th} simulation are as follows.

1. A dataset was generated with the above, fixed initialization parameters.

2. Particle filter was run with 1000 particles for 100 time steps. Figs. 2.1-2.4 give the tracking performance of the PF for one of the simulations; the solid curves denote the actual state and the dashed curves denote the PF.

Figure 2.3: Comparison between true state and particle filters estimate for the third state.



3. Two different quantities were calculated for each state:

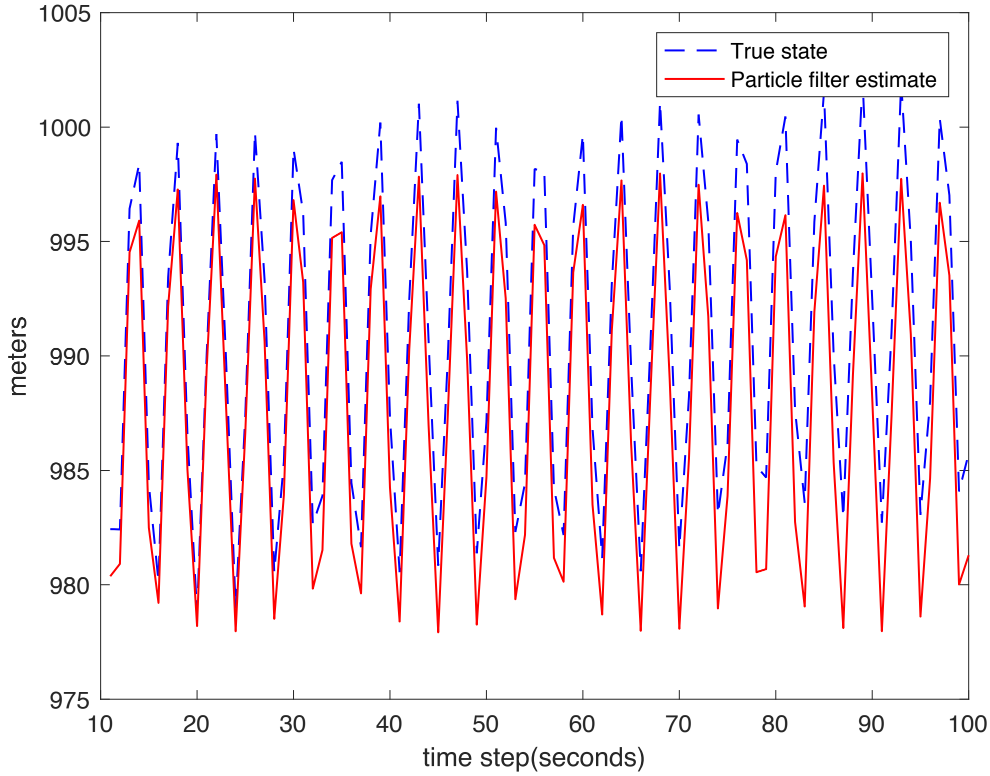
$$RMSE(s) = \sqrt{\frac{1}{100}\sum_{k=1}^{100}(\mathcal{X}^{(s)}(k) - \hat{\mathcal{X}}^{(s)}(k|k))^2}, \qquad s = 1, ..., 50.$$

and

$$SE_t(k, s) = (\mathcal{X}^{(s)}(k) - \hat{\mathcal{X}}^{(s)}(k|k))^2.$$

The first one is the root mean squared error averaged along the entire sample path for each simulation, whereas $SE_t(k, s)$ is squared error at each time step k, for simulation s.

Figure 2.4: Comparison between true state and particle filters estimate for the fourth state.



After running 50 simulations, the following two quantities are computed for each state:

$$AvRMSE = \frac{1}{50}\sum_{s=1}^{50} RMSE(s).$$

and

$$RMSE_t(k) = \sqrt{\frac{1}{50}\sum_{s} SE_t(s)}.$$

The results for AvRMSE with each of the state are summarized in Table 2.1. Figs. 2.5-2.8 give the $RMSE_t$ all the four states. It can be seen that the particle filter tracks the states of this highly nonlinear system quite well.

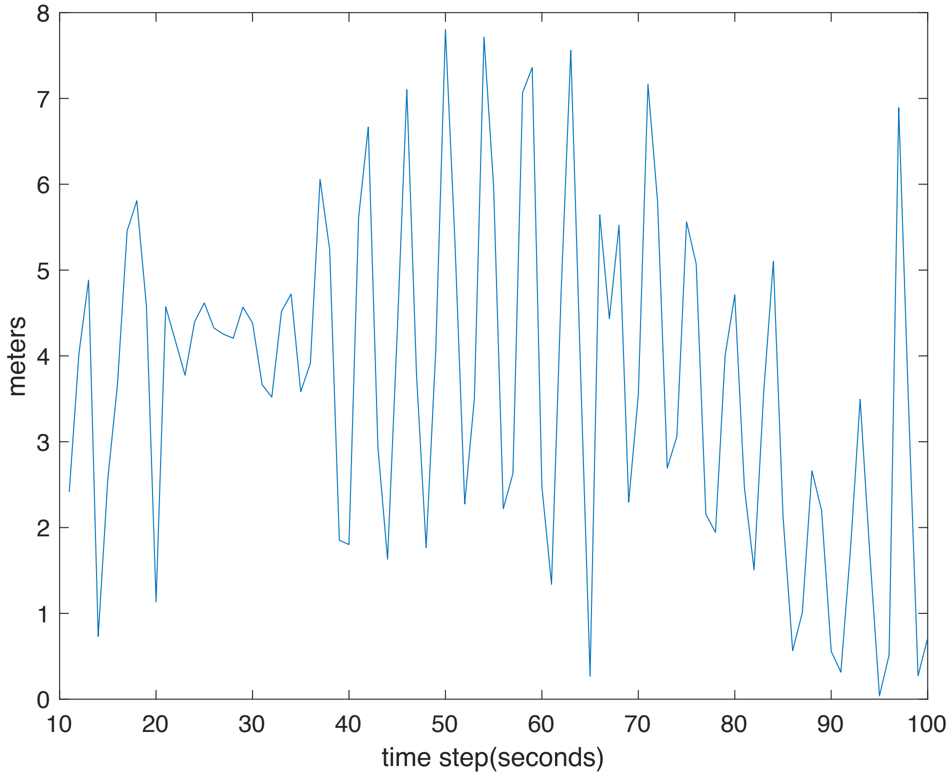
Table 2.1: $AvRMSE$ for each of the state

	$\mathcal{X}$	$\mathcal{Y}$	$\dot{\mathcal{X}}$	$\dot{\mathcal{Y}}$
$AvRMSE$	14.2715	8.7535	13.8102	8.5276

2.5 Summary

For a particular problem, if the assumptions of the Kalman filter hold, the Kalman filter will typically outperform other algorithms in terms of filtering performance. However, in many applications, the assumptions do not hold, and approximation techniques must be used. PFs and related sequential Monte Carlo methods provide an alternative class of algorithms for non-Gaussian and/or nonlinear system which approximates the density directly as a finite number of samples. A review of particle methods for filtering is provided. One application was presented to illustrate the performance of particle filtering algorithms for maneuvering target tracking. However, in some applications, it is still desirable to use a linearised filter instead of a PF for a variety of reasons, which may include hardware limitations in terms of computational time available for filtering at each time step. In the rest of the thesis, we focus on filters for linear state space systems with additive-multiplicative noise. In particular, the emphasis will be on dealing with randomly delayed measurements. We also look at the use of multiplicative noise as a proxy for the linearisation error. Note that the techniques developed in chapters 4 and 5 can also be used to generate a proposal density for a PF designed to deal with systems having random delays and additive-multiplicative noise, besides being standalone filters which are optimal in a certain sense in their own right. This strand of work is not developed further

Figure 2.5: $RMSE_t$ for the first state.



here , but has promise for future research.

Figure 2.6: $RMSE_t$ for the second state.

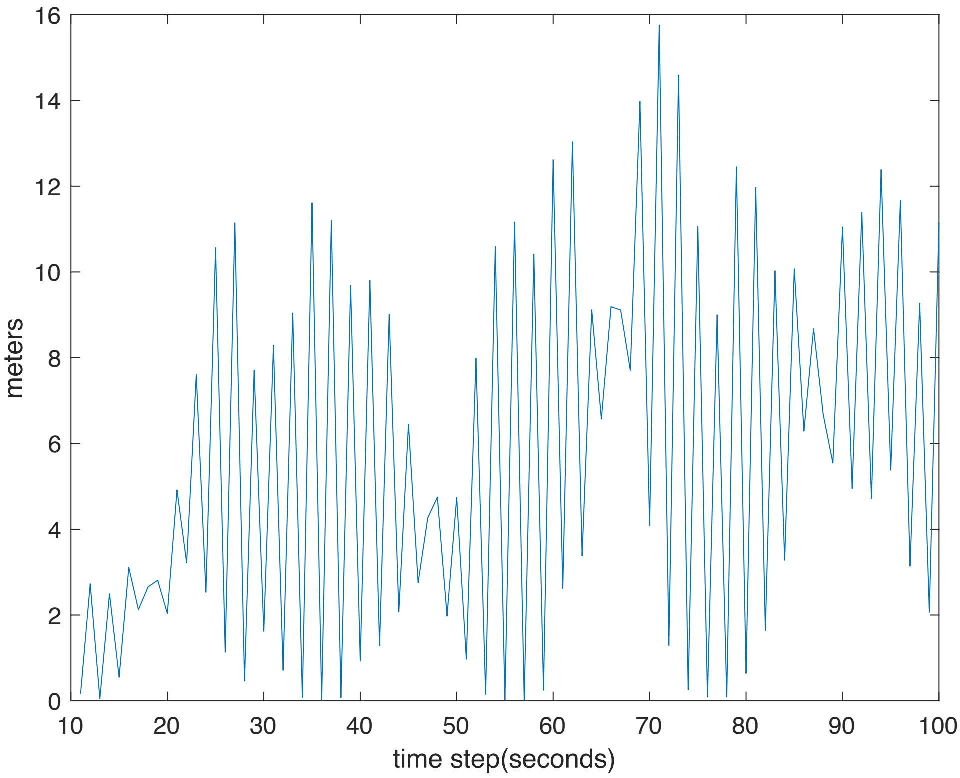


Figure 2.7: $RMSE_t$ for the third state.

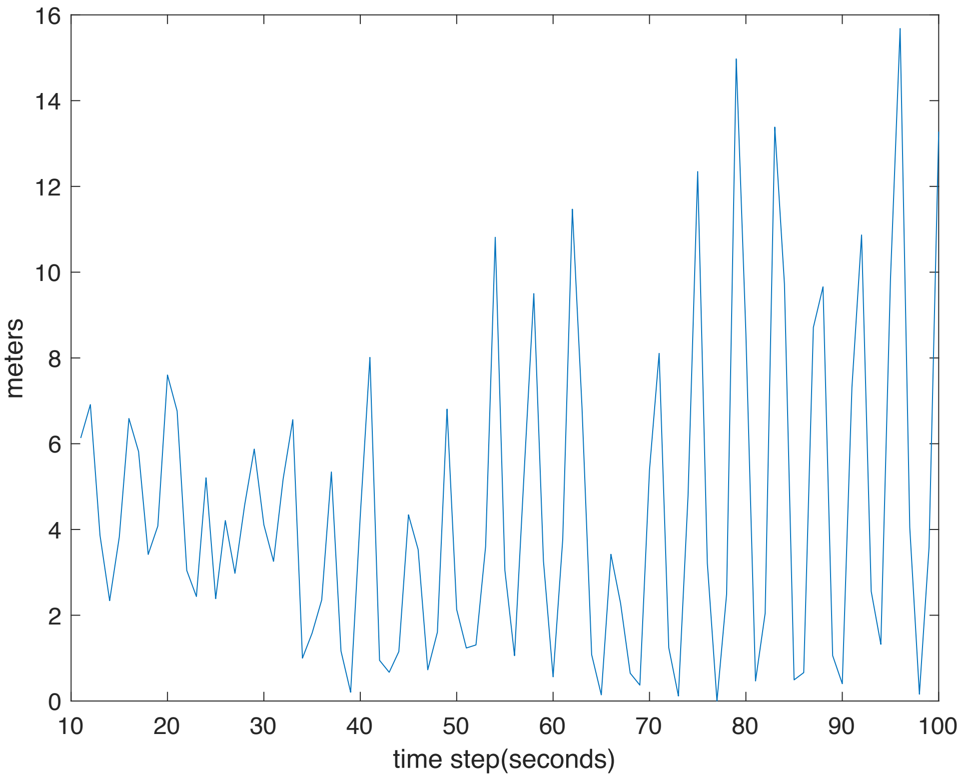
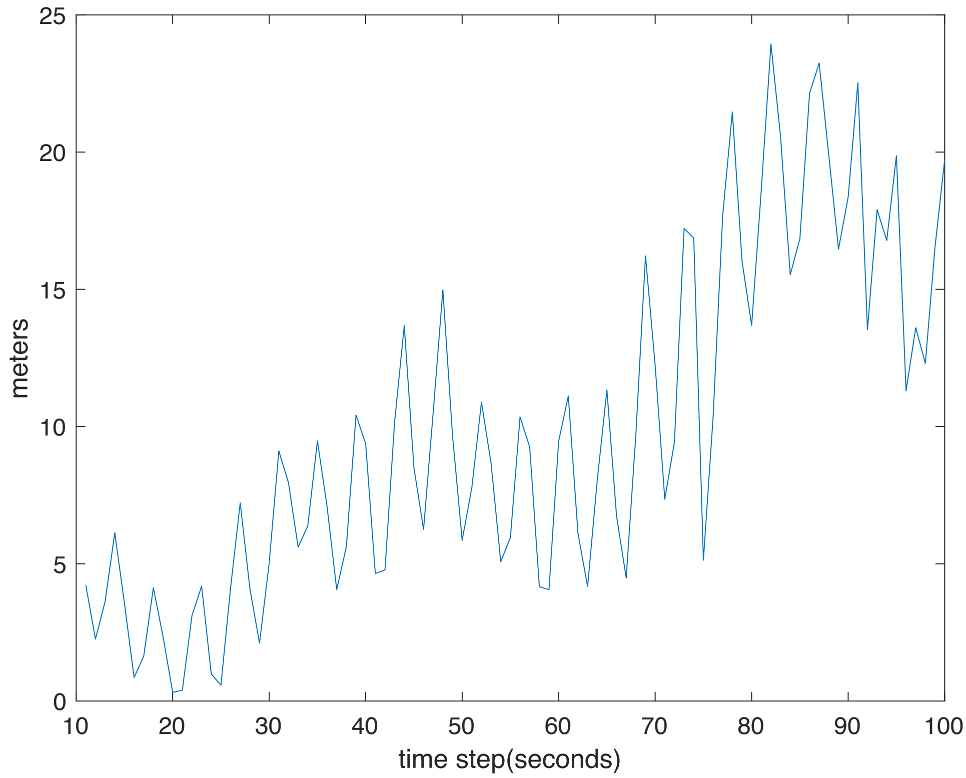


Figure 2.8: $RMSE_t$ for the fourth state.



Chapter 3

Approximate minimum variance filter under random delays I: linear discrete systems with additive and multiplicative noise

3.1 Introduction

In this chapter, we start with the problem of latent state estimation for linear discrete systems with additive-multiplicative noise. The main purpose of this chapter however, is the development of a new algorithm for discrete time linear systems with randomly delayed observations and additive-multiplicative noise. The material presented in this chapter has been published in [56] and submitted for publication in [57].

3.2 A minimum variance filter for discrete time linear systems with parametric uncertainty

3.2.1 Background

Estimation and filtering in noise is an important problem in many practical areas, including economics, tracking, weather forecasting, navigation systems, control and signal processing. The additive noise state space model has received much attention in the filtering literature (see, for instance, [1], [10], [11] and [58]). On the other hand, in many practical systems we need to consider the noise component to be both multiplicative and additive to the signal component. Compared to the additive noise case, the corresponding filtering problem for systems with multiplicative noise has received somewhat less attention. Multiplicative noise has been observed in many applications in the sciences and engineering, such as signal processing systems, chemistry, economics, biological movement and ecology; see [59], [60], [61], [62] and the references therein. A separate strand of research on the filtering problem for linear discrete time systems with multiplicative noise has also received a great deal of attention recently, since this kind of formulation has found many applications in the sciences and engineering. Examples of such systems are encountered in Doppler radar signal processing (see [63]), speech processing in signal-dependent noise [64], optical imaging under speckle or scintillation condition [65], sonar [66], synthetic aperture radar image processing [67], transmission of signals over fading channels [68] and mechanical vibrations [69] and [70]), where the multiplicative noise is mainly caused by nonlinearities in the observed system. In fact, [71] and [72] advocate using norm-bounded multiplicative uncertainty to model linearisation errors. The second-order statistics of the multiplicative noise, in contrast to the case of the additive noise, is unknown *a priori*, as it depends on the real state of the system. However, a multiplicative noise model can be used to model the stochastic uncertainty in the system parameters which are estimated from data, and a demonstration of this fact is one of the contributions in the chapter. In an EKF, multiplicative noise can act as a proxy for neglected higher-order terms in Taylor series (since, unlike additive noise, it does depend on the state). For systems subject to multiplicative noise, different kinds of methods have been introduced for the discrete time models. This includes a filtering algorithm which uses linear matrix inequalities to guarantee that the covariance error is bounded from above by a specified positive definite matrix (see [73]), an optimal filter within a class of polynomial transformations of a fixed degree (see [74]), a linear minimum mean square estimator (LMMSE) subject to state and measurement multiplicative noises and Markov jumps in the parameter vector (see [75]), a robust Kalman filter [76] and a MV linear filter for a class of systems which includes multiplicative noise (see [77]). In [78], a new structure of a linear recursive estimator minimizing the mean square error is derived for a system with multiplicative noise in the measurement model. The results in [78] were generalized in [79] to develop a different structure of a linear recursive estimator for a nonzero mean signal corrupted by multiplicative noise. In [59] the filtering and control problem under the H_∞ criterion is studied. In [80], a robust finite-horizon Kalman filter for discrete time varying uncertain system is designed with both deterministic uncertainties and stochastic uncertainties, with stochastic uncertainties expressed as multiplicative noise.

In the following subsections we consider an extension of Ponomareva and Date's work in [77] on filtering in systems with multiplicative noise. Specifically, we extend the results from [77] to propose a complete, closed-form solution to the MV filtering problem for linear systems with multiplicative noise in both transition and measurement equations and demonstrate its performance through numerical simulation experiments. One of the main contributions of the following subsections is a demonstration of how filtering under multiplicative noise can act as a proxy for filtering under parameter uncertainty, which is characterized as random perturbations of the state space matrices. A limited amount of work has been done on MV filtering under parametric uncertainty, although a related robust estimation problem under parametric uncertainty has recently been addressed in [81]. In the next subsection a MV filter for discrete time linear systems with parametric uncertainty is proposed.

3.2.2 System model and problem formulation

The system dynamics under consideration can be described by the following discrete time equation:

$$\mathcal{X}(k+1) = A\mathcal{X}(k) + B + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k), \tag{3.1}$$

while the measurement model is

$$\mathcal{Y}(k) = C\mathcal{X}(k) + D + U_v\mathcal{V}(k) + G_2 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_2(k). \tag{3.2}$$

Here $\mathcal{X}(k) \in \mathbb{R}^n$ is the state vector at time k, $\mathcal{Y}(k) \in \mathbb{R}^r$ is the measurement vector at time k and A, B, G_1, C, D, G_2, U_w and U_v are given deterministic matrices. For a vector $\mathcal{Z}, M = \operatorname{diag}(\mathcal{Z})$ represents a diagonal matrix with $M_{jj} = \mathcal{Z}_j$. $\mathcal{W}(k) \in \mathbb{R}^n$ and $\mathcal{V}(k) \in \mathbb{R}^r$ are the process noise and the measurement noise, respectively. The random variables $\mathcal{S}_1(k) \in \mathbb{R}^n$ and $\mathcal{S}_2(k) \in \mathbb{R}^r$ represent the multiplicative noise sources. Note that (3.1) and (3.2) contain both additive and multiplicative noise. We make two assumptions:

1. The noise signals $\mathcal{W}(k)$, $\mathcal{V}(k)$, $\mathcal{S}_1(k)$ and $\mathcal{S}_2(k)$ are zero mean, i.i.d. random vectors with covariance matrix
   $\mathbb{E}[\mathcal{V}(k)\mathcal{V}^\top(j)] = \{ \begin{matrix} 0 & if & k \neq j, \\ I & if & k = j, \end{matrix}$
   $\mathbb{E}[\mathcal{W}(k)\mathcal{W}^\top(j)] = \{ \begin{matrix} 0 & if & k \neq j, \\ I & if & k = j, \end{matrix}$
   and $\mathbb{E}[\mathcal{S}_i(k)\mathcal{S}_i^\top(j)] = \{ \begin{matrix} 0 & if & k \neq j & i = 1, 2, \\ I & if & k = j & i = 1, 2, \end{matrix}$
   and are uncorrelated with each other.

2. The initial state is a random vector with a known mean and covariance matrix, $\mathbb{E}[\mathcal{X}(0)] = \hat{\mathcal{X}}(0)$ and $\mathbb{E}[(\mathcal{X}(0) - \hat{\mathcal{X}}(0))(\mathcal{X}(0) - \hat{\mathcal{X}}(0))^\top] = \bar{P}(0)$, respectively. $\mathcal{X}(0)$, $\mathcal{W}(k)$, $\mathcal{V}(k)$, $\mathcal{S}_1(k)$ and $\mathcal{S}_2(k)$ are mutually independent.

Under the above assumptions, note that one may also treat $\mathcal{S}_1(k)$ and $\mathcal{S}_2(k)$ as zero mean random perturbations in the system matrices A and C, which have the covariance matrices $G_1G_1^\top$ and $G_2G_2^\top$ respectively. This can be seen by re-arranging (3.1) and (3.2) as

$$\mathcal{X}(k+1) = (A + G_1\text{diag}(\mathcal{S}_1(k)))\,\mathcal{X}(k) + B + U_w\mathcal{W}(k), \tag{3.3}$$

$$\mathcal{Y}(k) = (C + G_2\text{diag}(\mathcal{S}_2(k)))\,\mathcal{X}(k) + D + U_v\mathcal{V}(k). \tag{3.4}$$

Later, we use this interpretation of multiplicative noise as stochastic perturbations in parameters in one of the numerical experiments in section 4.2.3.

Remark 1. Since the system (3.1)-(3.2) contains multiplicative noise, it is non-Gaussian. The exact conditional mean (which minimizes the conditional variance) is no linear in the past measurement. Hence our filtering algorithm (which minimizes the conditional variance over the set of filters which are linear in the current measurement) is an *approximate* MV filter.

Assume that the observations up to time k are given and that the conditional mean of $\mathcal{X}(k)$ given $\mathcal{Y}(k)$, $\hat{\mathcal{X}}(k|k)$, is available. From this value, the approximated conditional mean of $\mathcal{X}(k+1)$, which provides the predictor $\hat{\mathcal{X}}(k+1|k)$, is derived using (3.1):

$$\hat{\mathcal{X}}(k+1|k) = A\hat{\mathcal{X}}(k|k) + B. \tag{3.5}$$

The predictor $\hat{\mathcal{X}}(k+1|k)$ must now be updated with the information provided by the new measurement $\mathcal{Y}(k+1)$ to obtain the filter. The update equation for a linear filter is

$$\hat{\mathcal{X}}(k+1|k+1) = \hat{\mathcal{X}}(k+1|k) + \bar{K}(k+1)(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)), \tag{3.6}$$

where $\bar{K}(k+1)$ is the filter gain and $\hat{\mathcal{Y}}(k+1|k)$ is a one step ahead prediction of $\mathcal{Y}(k+1)$. The estimation error covariance matrix is given by

$$\bar{P}(k+1|k+1) = \mathbb{E}[\Phi(k+1)\Phi(k+1)^\top], \tag{3.7}$$

where $\Phi(k) := \mathcal{X}(k+1) - \hat{\mathcal{X}}(k+1|k+1)$. Before presenting the main result in this section, we first introduce the following useful lemma. In the subsequent discussion, $vec(f(x_j))$ denotes a vector with $f(x_j)$ as its j^{th} element.

Lemma 1. The second order moment of $\operatorname{diag}(\mathcal{X}(k))$ is given as follows:

$$q(k|k) = \operatorname{diag}\left(vec(P_{jj}(k|k)) + (\hat{\mathcal{X}}(k|k))^2\right), \tag{3.8}$$

where $q(k|k) = \mathbb{E}[\operatorname{diag}(\mathcal{X}(k))\operatorname{diag}(\mathcal{X}(k))^\top]$.

Proof

$$q(k|k) = \mathbb{E}[\operatorname{diag}(\mathcal{X}(k))\operatorname{diag}(\mathcal{X}(k))^\top] = \mathbb{E}[(\operatorname{diag}\mathcal{X}(k))^2]. \tag{3.9}$$

On the other hand, based on the definition of the covariance matrix we have

$$\begin{aligned}
&\operatorname{diag}\left(vec(\bar{P}_{jj}(k|k))\right) = \mathbb{E}[(\operatorname{diag}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))^2] \\
&= \mathbb{E}[(\operatorname{diag}\mathcal{X}(k))^2] - 2\mathbb{E}[\operatorname{diag}\mathcal{X}(k)]\operatorname{diag}\hat{\mathcal{X}}(k|k) + (\operatorname{diag}(\hat{\mathcal{X}}(k|k)))^2 \\
&= \mathbb{E}[(\operatorname{diag}(\mathcal{X}(k)))^2] - 2\operatorname{diag}\hat{\mathcal{X}}(k|k)\operatorname{diag}\hat{\mathcal{X}}(k|k) + (\operatorname{diag}(\hat{\mathcal{X}}(k|k)))^2 \\
&= \mathbb{E}[(\operatorname{diag}(\mathcal{X}(k)))^2] - (\operatorname{diag}(\hat{\mathcal{X}}(k|k)))^2,
\end{aligned} \tag{3.10}$$

then

$$\mathbb{E}[(\operatorname{diag}\mathcal{X}(k))^2] = \operatorname{diag}\left(vec(P_{jj}(k|k)) + (\hat{\mathcal{X}}(k|k))^2\right), \tag{3.11}$$

and by substituting (3.11) in (3.9), the proof of (3.8) can be completed.

Our objective is to find a filter gain $\bar{K}(k+1)$ that would minimize the trace of the estimation error covariance matrix $\bar{P}(k+1|k+1)$ of the state estimate $\hat{\mathcal{X}}(k+1|k+1)$ and obtain an expression for the filter. Our main result in this section, which is an extension of the corresponding result from [77], is given in the next theorem.

Theorem 1. For system (3.1)-(3.2) with assumptions 1-2, the filter gain $\bar{K}(k+1)$ that minimizes the trace of the covariance $\bar{P}(k+1|k+1)$ is given by

$$\bar{K}(k+1) = \bar{P}(k+1|k)C^\top[C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k|k)G_2^\top]^{-1}, \tag{3.12}$$

where $q(k|k)$ is as defined in equation (3.8) and

$$\bar{P}(k+1|k) = A\bar{P}(k|k)A^\top + U_wU_w^\top + G_1q(k|k)G_1^\top, \tag{3.13}$$

$$\begin{aligned}&\bar{P}(k|k) = \bar{P}(k|k-1) + \bar{K}(k)(C\bar{P}(k|k-1)C^\top + U_vU_v^\top + \\ &G_2q(k|k-1)G_2^\top)\bar{K}(k)^\top - \bar{P}(k|k-1)C^\top\bar{K}(k)^\top - \bar{K}(k)C\bar{P}(k|k-1),\end{aligned} \tag{3.14}$$

Proof: See Appendix.
Several comments on this result are in order.

- $\bar{P}(k+1|k)$ in (3.13) is linear in $q(k|k)$, which in turn is nonlinear in $\mathcal{Y}(k)$. This means that $\bar{K}(k+1)$ is a nonlinear function of $\mathcal{Y}(k)$.

- If $G_2 = 0$, our filter reduces to a special case of the filter with state multiplicative noise only, which has previously been discussed in [77], with $\gamma = 1$ in the authors' notation in that paper. If, in addition, $G_1 = 0$ *i.e.,* if there is no multiplicative noise either in the transition equation or in the measurement equation, our filter reduces to the Kalman filter for the linear additive noise case.

- Note that, under fairly general conditions, the Kalman filter described in chapter 2 (equations (2.11)-(2.17)) converges to a constant gain matrix as $k \to \infty$ (see, e.g. [1]). In the case of bilinear systems, no such convergence results are available (see, e.g. [71]-[80]). In fact, noting as above that $\bar{P}(k+1|k)$ is a function of $q(k)$ and hence of $\mathcal{Y}(k)$, $\bar{K}(k+1)$ cannot converge to a time-invariant matrix.

3.2.3 Numerical examples

We consider two numerical examples in this section. In both cases, the parameters of the state space model under consideration are estimated from real data in the literature. The purpose of these examples is to demonstrate how one can use the information about randomness in the parameters estimated from data (*e.g.*, in terms of the asymptotic variance of the parameter estimates) to design a filter which minimizes the variance relative to the additive noise as well as the *noise* introduced by parameter uncertainty.

Example 1

In this example, we consider a two-factor extension of the Vasicek interest rate model [82]. The treatment below and the estimated parameter values are from [83], although the model has been treated quite extensively in the econometric and financial literature; see, e.g., [84] and [85] for empirical studies, among others. The key assumption of the two-factor Vasicek model is that instantaneously compounded interest rate (or the short rate) given by the sum of two state variables, each of them following an Ornstein-Uhlenbeck process. Let us consider two independent state variables that follow linear, mean reverting Gaussian process under the risk neutral measure $\mathcal{Q}$

$$\begin{aligned} &\mathcal{X}(t) = \mathcal{X}_1(t) + \mathcal{X}_2(t), \\ &d\mathcal{X}_i(t) = \kappa_i(\theta_i - \mathcal{X}_i(t))dt + \sigma_i d\mathcal{W}_i(t) \quad \mathcal{X}_i(0) = \mathcal{X}_{i0} \quad i = 1, 2, \end{aligned} \tag{3.15}$$

where $\mathcal{X}_{i0}$, κ_i, θ_i and σ_i are positive constants, and $\mathcal{W}_i(k)$ are uncorrelated $\mathcal{Q}$-Wiener processes. Each $\mathcal{X}_i(k)$ conditional to $\mathcal{F}_s$ is normally distributed with mean and variance:

$$\mathbb{E}[\mathcal{X}_i(t)|\mathcal{F}_s] = \mathcal{X}_i(s)e^{-\kappa_i(t-s)} + \theta_i(1 - e^{-\kappa_i(t-s)}) \tag{3.16}$$

$$Var[\mathcal{X}_i(t)|\mathcal{F}_s] = \frac{\sigma_i^2}{2\kappa_i}(1 - e^{-2\kappa_i(t-s)}) \tag{3.17}$$

We discretize the two equations (3.16) and (3.17) considering evenly spaced observation times $t_1 \leq t_2 \leq \cdots \leq t_N$ where $t_{n+1} - t_n = \Delta t$. So, the transition equation in the state space formulation of this model is expressed as follows:

$$\begin{bmatrix} \mathcal{X}_1(k+1) \\ \mathcal{X}_2(k+1) \end{bmatrix} = \underbrace{\begin{bmatrix} e^{-\kappa_1 \Delta t} & 0 \\ 0 & e^{-\kappa_2 \Delta t} \end{bmatrix}}_{A} \begin{bmatrix} \mathcal{X}_1(k) \\ \mathcal{X}_2(k) \end{bmatrix} + \begin{bmatrix} \theta_1(1 - e^{-\kappa_1 \Delta t}) \\ \theta_2(1 - e^{-\kappa_2 \Delta t}) \end{bmatrix} + \begin{bmatrix} \mathcal{W}_1(k+1) \\ \mathcal{W}_2(k+1) \end{bmatrix}, \tag{3.18}$$

where $\mathcal{W}(k+1) = \begin{bmatrix} \mathcal{W}_1(k+1) & \mathcal{W}_2(k+1) \end{bmatrix}^\top \sim \mathcal{N}(0, W)$, with

$$W = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1}(1 - e^{-2\kappa_1 \Delta t}) & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2}(1 - e^{-2\kappa_2 \Delta t}) \end{bmatrix},$$

This discretisation preserves the exact conditional mean $\mathbb{E}[\mathcal{X}_i(t)|\mathcal{F}_t]$ and the exact conditional variance $Var[\mathcal{X}_i(t)|\mathcal{F}_s]$. When the sort rate follows the stochastic process given by (3.14), zero coupon bond price at time t, for a bond which matures at time $T > t$ is given in the following analytical form:

$$P(t, T, \mathcal{X}_1(t), \mathcal{X}_2(t)) = \mathbb{E}(e^{-\int_t^T \mathcal{X}_s ds}) = e^{E(t,T) - F_1(t,T)\mathcal{X}_1(t) - F_2(t,T)\mathcal{X}_2(t)},$$

where

$$E(t, T) = \sum_{i=1}^{2} \frac{(\kappa_i^2(\theta_i - \frac{\sigma_i \lambda_i}{\kappa_i}) - \frac{\sigma_i^2}{2})(F_i(t, T) - (T - t))}{\kappa_i^2} - \frac{\sigma_i^2 F_i^2(t, T)}{4\kappa_i},$$

and

$$F_i(t, T) = \frac{1}{\kappa_i}(1 - e^{-\kappa_i(T-t)}), \quad i = 1, 2.$$

where λ_i is the market price of risk for the i^{th} factor. The measurement system we used involves the following relationship between zero-coupon yields and the price of zero-coupon bonds:

$$\mathcal{Y}(t, T) = \frac{-E(t, T) + \sum_{i=1}^{2} F_i(t, T)\mathcal{X}_i(t)}{T - t} \tag{3.19}$$

In other words,

$$P(t,T,\mathcal{X}_1(t),\mathcal{X}_2(t)) = \exp(-y(t,T)(T-t)).$$

Using equation (3.19) at each t_n, for a set of m bonds with maturities $T_1,...,T_m$ leads to the following vector valued equation:

$$\begin{bmatrix} \mathcal{Y}(t_k,T_1) \\ \mathcal{Y}(t_k,T_2) \\ . \\ . \\ . \\ \mathcal{Y}_1(t_k,T_m) \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{F_1(t_k,T_1)}{T_1-t_k} & \frac{F_2(t_k,T_1)}{T_1-t_k} \\ \frac{F_1(t_k,T_2)}{T_2-t_k} & \frac{F_2(t_k,T_2)}{T_2-t_k} \\ . & . \\ . & . \\ . & . \\ \frac{F_1(t_k,T_m)}{T_m-t_k} & \frac{F_2(t_k,T_m)}{T_m-t_k} \end{bmatrix}}_{C} \begin{bmatrix} \mathcal{X}_1(k) \\ \mathcal{X}_2(k) \end{bmatrix}$$

$$+ \begin{bmatrix} -\frac{E_1(t_k,T_1)}{T_1-t_k} \\ -\frac{E_1(t_k,T_2)}{T_2-t_k} \\ . \\ . \\ . \\ -\frac{E_1(t_k,T_m)}{T_m-t_k} \end{bmatrix} + \begin{bmatrix} \mathcal{V}_1(k) \\ \mathcal{V}_2(k) \\ . \\ . \\ . \\ \mathcal{V}_m(k) \end{bmatrix},$$

Here, κ_i, σ_i, λ_i and θ_i, $i = 1, 2$ are model parameters. For the example from [83] considered here, $m = 6$. $\mathcal{V}_i(k) \sim \mathcal{N}(0, H)$ are noise variables with $H = diag(h_1^2, h_2^2, ..., h_m^2)$, where h_i are positive constants. In practice, equation (3.18) is the two-factor short rate model, while each $\mathcal{Y}(t_k, T_i)$ denotes the yield at time t_k for maturity T_i. The parameters used in our simulation are the same as those estimated from the real data in [83]. These are listed in Table 3.1; see [83] for the exact details of the data set, parameter estimation procedure, etc. As the parameters are obtained from data, one has an estimate of the standard error in each parameter. We consider a small, zero mean, normally distributed perturbation in the nominal values of κ_1 and κ_2, with a standard deviation of $\beta = 5\%$ of the nominal value of each of the two parameters. Matrices G_1 and G_2 are then computed to reflect the element-wise uncertainties introduced in matrices A and C respectively, using Monte Carlo simulation. This experiment is then repeated for changed standard deviations of $\beta = 10\%$ and $\beta = 15\%$ of the nominal values for both parameters.

The initial conditions as in [83] are used:
$\mathcal{X}(0) = \begin{bmatrix} 0.015 & 0.025 \end{bmatrix}^\top$, $\hat{\mathcal{X}}(0) = \begin{bmatrix} 0.02 & 0.02 \end{bmatrix}^\top$ and $P(0) = 5\times 10^{-3} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Table 3.1: Estimated parameters for Vasicek model [83]

Parameters	κ_1	θ_1	σ_1	λ_1	κ_2
Values	0.7030	0.0056	0.0321	-0.4591	0.0255
Parameters	θ_2	σ_2	λ_2	h_1	h_2
Values	0.0035	0.0142	-0.2652	0.0009	0.0012
Parameters	h_3	h_4	h_5	h_6	
Values	0.0013	0.0007	0.0009	0.0010	

Our goal in this section is to see whether a small uncertainty or a small random perturbation in the parameter values has an impact on the filter ring performance. We will compare the Kalman filter (KF) (which ignores the parameter uncertainty) and the filter in [77] (which ignores the measurement multiplicative noise) with the filter proposed in this section (called MKF in the tables in this section), which encapsulates the parameter uncertainty in terms of multiplicative noise via G_1 and G_2 matrices. In order to compare the performance of the estimators, we use the root mean square error (RMSE) criterion. Consider 100 independent simulations, each with 200 data points. Denoting $\mathcal{X}^{(s)}(k)$, $\quad k = 1, ..., 200$ as the s^{th} set of true values of the state, and $\hat{\mathcal{X}}^{(s)}(k|k)$ as the filtered state estimate at time k for the s^{th} simulation run, the RMSE of the filter for each of the algorithms is calculated by

$$RMSE_i(s) = \sqrt{\frac{1}{200}\sum_{k=1}^{200}(\mathcal{X}_i^{(s)}(k) - \hat{\mathcal{X}}_i^{(s)}(k|k))^2},$$
$$i = 1, 2, \qquad s = 1, ..., 100.$$

Then, the average of RMSE for each of the states over 100 simulations is given by

$$AvRMSE_i = \frac{1}{100}\sum_{s=1}^{100} RMSE_i(s), \quad i = 1, 2.$$

The results for the three different levels of perturbations, $\beta = 5\%$, $\beta = 10\%$ and $\beta = 15\%$, are summarised in Table 3.2. Recall that each of the parameters κ_1 and κ_2 are perturbed by normally distributed random noise with zero mean and a standard deviation equal to β times their respective nominal values to generate G_1, G_2 matrices in the state space equations using Monte Carlo simulation. As can be seen, the modified filter, *i.e.* the MKF has a significantly smaller AvRMSE than the KF and the filter in [77] for both states and for all three levels of parameter perturbations, since the parameter uncertainties are not taken into account in the KF and the measurement multiplicative noise is not taken into account in [77].

Example 2

As another example with parameters estimated from real data, consider a discrete-time system (3.1)-(3.2) with the following parameter specification:

$$A = \rho \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ \sin(\lambda) & \cos(\lambda) \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad B = 0, \quad D = 0$$

where $U_w^2 = 214$ and $U_v^2 = 1593$. ρ and λ are random variables with means $\hat{\rho} = 0.4$ and $\hat{\lambda} = 0.41$. The dynamics and the chosen parameter values (with $G_1 = 0$ and $G_2 = 0$) are used in ([86], chapter 2) as a time series model

Table 3.2: Comparison of $AvRMSE_1$ and $AvRMSE_2$ for KF, filter in [77] and MKF for different levels of perturbations β

Level of perturbations		$\beta = 0.05$	$\beta = 0.10$	$\beta = 0.15$
$AvRMSE_1 \times 10^{-2}$	KF	137.0532	137.0534	137.0536
	Filter in [77]	136.9530	136.9531	136.9533
	MKF	136.4515	135.9493	135.3313
$AvRMSE_2 \times 10^{-2}$	KF	108.1107	108.1108	108.1109
	Filter in [77]	107.9105	107.9106	107.9107
	MKF	107.5329	107.0198	106.9983

of rainfall in north-east Brazil. As in the previous subsection, we allocate normally distributed random perturbations to λ and ρ with zero mean and β times the nominal values as the standard deviation. Three different values of β are used, as in the previous case: $\beta = 5\%$, $\beta = 10\%$ and $\beta = 15\%$. The perturbation matrix G_1 is computed in each case via Monte Carlo simulation. G_2 in this case is zero since there is no uncertainty in $C = \begin{bmatrix} 1 & 1 \end{bmatrix}$. The initial conditions are
$\mathcal{X}(0) = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$, $\quad \hat{\mathcal{X}}(0) = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$ and $P(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Similar to the previous subsection, the RMSE of the filter for each of the two algorithms is calculated by

$$RMSE(s) = \sqrt{\frac{1}{200}\sum_{k=1}^{200}(\sum_{i=1}^{2}\mathcal{X}_i^{(s)}(k) - \sum_{i=1}^{2}\hat{\mathcal{X}}_i^{(s)}(k|k))^2}, \quad s = 1, ..., 100.$$

Note that the error in this case is the difference between the true rainfall generated by the model and the filtered estimate of rainfall. The average of

RMSE over 100 simulations is given by

$$AvRMSE = \frac{1}{100}\sum_{s=1}^{100} RMSE(s).$$

Table 3.3 summarises the results of this numerical experiments. As can be seen, the proposed filtering algorithm clearly outperforms the KF, which ignores the parameter uncertainties. This is in keeping with the theoretical results in [77].

Table 3.3: Comparison of $AvRMSE$ for KF and MKF for different levels of perturbations β

Level of perturbations		$\beta = 0.05$	$\beta = 0.10$	$\beta = 0.15$
$AvRMSE$	KF	6.3097	6.3697	6.5014
	MKF	5.8086	5.4468	5.0106

3.3 A minimum variance filter for discrete time linear systems with randomly delayed observations and additive and multiplicative noise

3.3.1 Background

Most traditional discrete time filter design approaches depend on the assumption that the measurements generated by the system at each time step contain information about the state of the system at that time step. However, random delays may affect the arrival of measurements in various practical situations. This occurs widely, for instance, when a sensor is connected to the estimator through a network with limited bandwidth. Let us consider that an unmanned air vehicle (UAV) is controlled by an operator through a communication network with a remote sensor. The operator should know the UAV's state in order to control its motion. However, if the information sent by the sensor reaches the operating terminal after a random time delay, the location of the UAV may be incorrectly estimated and this may lead to a hazardous situation. This example demonstrates the need for dealing specifically with random delays when designing a filter.

In networked systems as well, due to their limited carrying capacity and bandwidth, unavoidable uncertainties of communication networks such as packet dropouts, random delays and missing measurements estimation performance may be affected. Accordingly, filtering problems for network-induced phenomena such as packet dropouts, measurements delays and missing measurements have been proposed. For systems with packet dropouts, a large number of filter design algorithms have been investigated in the literature (see, e.g., [87], [88], [89], [90], [91] and references therein). Some system measurements may contain noise only at certain time points so that the true signals are simply missing. As such the filtering problems with missing measurements have received considerable research attention and many important results have been reported in recent years (see, e.g., [92], [93], [94], [95] and [96]).

In recent years, several studies have focused on the estimation of the state of discrete-time linear systems with randomly delayed observations (see, e.g.,[97] and [98]). These models have received considerable research attention, and many important results have been reported in the literature. Linear filtering for discrete time systems with finite random measurement delays is investigated in [99]. In [100], an optimal filtering problem for networked systems with random transmission delays is investigated using a multi-state Markov chain model for the delay process. Centralized fusion filters are designed in [101] for linear systems with multiple sensors with different delay rates. In [102], a modification of the minimum variance state estimator is employed to accommodate the effects of random delays in sensor data arrival at the controller terminal. Robust filtering problems with random delays are investigated in [103]. The linear unbiased estimator is studied in [104], where the one-step sensor delay is described as a binary white noise sequence. Furthermore, many researchers have investigated these problems under different assumptions about the possible delays and different filtering approximations. As an example, when the random delay was characterised by a set of Bernoulli variables, the linear quadratic filter and fixed-point smoother from randomly delayed observations have been proposed in [105, 106] using covariance information about the state and the observations. The finite horizon optimal filter, predictor and smoother are given by innovation analysis method [107]. The proposed estimators are derived by applying the innovation approach and do not require the knowledge of the state-space model; they use as information the second-order moments of the processes involved and the probability of delay in the observations. The filtering problem has been investigated when time delays are unknown in [108] by augmenting the system and then applying the Kalman filter.

However, multiplicative noise (or any other state-dependent noises) are not taken into account in the above papers and only additive noise is considered in the state transition equation. Multiplicative noise is often characterised as signal dependent. This kind of noise accounts for disturbances in known signals with unknown parameters, as well as random signals with Gaussian or non-Gaussian additive noise. However, the literature on filtering problems arising from delayed observations as well as multiplicative noise is far less extensive. In [109], the problem of a single random delay in each of the sensor measurements is studied. Multiple random delays may occur in many applications, and a study of multiple random delay systems under possibly parametric uncertainty in the state space models (modelled as multiplicative noise) is clearly worthy of further investigation. The optimal Kalman filter contaminated with multiplicative noises and randomly occurring two-step sensor delays is designed in [110]. Recently, the centralized and distributed fusion estimation problems for discrete-time system with random parameter matrices and subject to random delays and packet dropouts is investigated in [111]. The quantised filtering problem for a class of nonlinear time-varying systems subject to multiplicative noise and missing measurements is investigated in [112], the measurement multiplicative noise is also considered in this paper, and an upper bound on the covariance matrix is minimized by the designed filter. The optimal linear estimators for networked control system with uncertainties, multiple sensors and packet losses are derived in [113]. In [114], an innovation analysis approach is employed for filtering in a discrete-time stochastic system with multiple sensors subject to multiplicative noise and missing measurements. In [115], the MV estimation has been developed for linear discrete time-varying subject to bounded uncertainty when the packet are with and without time-stamp. Recently, the same problem is considered with correlated noises; see e.g., [116, 117, 118, 119] and [120].

Motivated by the above discussions, a new algorithm for approximate MV filtering in the presence of both additive and multiplicative noise, when the measurements are randomly delayed is proposed in this section. A multiplicative noise term is considered in both the process and the measurement equations to deal with stochastic uncertainty which arises out of either linearisation errors or parametric uncertainty. The observation delays are described by mutually independent random variables where the number of samples which the observation is delayed is uncertain. A complete closed-form solution to the delayed filtering problem for this class of systems is designed such that the filter gain minimizes the trace of covariance of the estimation error. The minimization is over a set of filters which have a Kalman filter-like structure, *i.e.*, the filter is linear in the current measurement, although it is a nonlinear function of past measurements. We demonstrate our results through numerical examples, that also extend our interpretation of multiplicative uncertainty as stochastic uncertainty in parameters, which was recently suggested in [56], to the case with random delays.

3.3.2 System model and problem formulation

The system dynamics under consideration can be described by (3.1) while the measurement model is as (3.2). In addition, A is assumed to be a full rank matrix.

One-step randomly delayed observation

To describe the situation when there is at most one random delay at each sample time, we use a sequence of independent Bernoulli random variables whose probabilities are known only to the model whether the measurement arriving at each sample time is delayed or not. At each measurement time k, there is a Bernoulli random variable p_k indicating whether the measurement is delayed or not. $p_k = 0$ and $p_k = 1$ represent the cases when the measurement is up to date and it is delayed by one sampling period, respectively. If the delay is not random and $\mathcal{Y}(k)$ depends on $\mathcal{X}(k-j)$ for a fixed j, note that the problem is much simpler.

It is assumed that, at $k = 1$, $\mathcal{Y}(1)$ is available for state estimation. At each time $k > 1$, the observation may either be delayed randomly by one sampling period or may be updated. The probability of update is denoted by β, so the probability of delay is $1 - \beta$. Thus the observation available at time $k > 1$ can be described as follows:

$$\begin{aligned} \mathcal{Z}(k) &= \mathcal{Y}(k), \quad \textit{with} \quad \textit{probability} \quad \beta, \\ \mathcal{Z}(k) &= \mathcal{Y}(k-1), \quad \textit{with} \quad \textit{probability} \quad (1-\beta). \end{aligned} \tag{3.20}$$

We further assume that the observation delay at sample time k is independent of all other sources of randomness in the system, viz, the delays at other sample times $j \neq k$, the initial state $\mathcal{X}(0)$ and other noise sources ($\mathcal{W}(k)$, $\mathcal{V}(k)$, $\mathcal{S}_1(k)$ and $\mathcal{S}_2(k)$).

The one step randomly delayed measurement equation (alternatively referred to as "received measurement") can now be written as:

$$
\begin{aligned}
&\mathcal{Z}(k) = (1-p_k)\mathcal{Y}(k) + p_k\mathcal{Y}(k-1),\\
&\mathbb{P}(p_k = 0) = \beta,\\
&\mathbb{P}(p_k = 1) = 1-\beta = \mathbb{E}(p_k),\\
&\mathbb{E}[(p_k-(1-\beta))^2] = \beta(1-\beta).
\end{aligned} \tag{3.21}
$$

$p_k = 0$ implies that the current measurement has arrived (i.e., the measurement is up to date), while $p_k = 1$ means that the measurement at the previous time step has arrived (*i.e.*, the measurement is delayed by one step). Here, $0 \le \beta \le 1$ represents the probability of no delay in measurement. Using (3.21) we can write our one step ahead prediction of $Z(k)$ in terms of the past predictions of $\hat{\mathcal{Y}}(k-i|k-1), \quad i = 0, 1$ as

$$
\hat{\mathcal{Z}}(k|k-1) = \beta\mathcal{Y}(k|\hat{k}-1) + (1-\beta)\hat{\mathcal{Y}}(k-1|k-1) \tag{3.22}
$$

Putting (3.1), (3.2) and (3.21) together, we have

$$
\begin{aligned}
\mathcal{X}(k+1) &= A\mathcal{X}(k) + B + U_w\mathcal{W}(k) + G_1\,\mathrm{diag}(\mathcal{X}(k))\mathcal{S}_1(k),\\
\mathcal{Y}(k) &= C\mathcal{X}(k) + D + U_v\mathcal{V}(k) + G_2\,\mathrm{diag}(\mathcal{X}(k))\mathcal{S}_2(k),\\
\mathcal{Z}(k) &= (1-p_k)\mathcal{Y}(k) + p_k\mathcal{Y}(k-1).
\end{aligned} \tag{3.23}
$$

To derive the recursive filtering equations, it is assumed that the observations are given up to time k and that the conditional mean of $\mathcal{X}(k)$ given $\mathcal{Z}(k)$, $\hat{\mathcal{X}}(k|k)$, is available. From this value, the conditional mean of $\mathcal{X}(k+1)$, which provides the predictor, $\hat{\mathcal{X}}(k+1|k)$, is derived using (3.1):

$$
\hat{\mathcal{X}}(k+1|k) = A\hat{\mathcal{X}}(k|k) + B. \tag{3.24}
$$

The predicted estimate $\hat{\mathcal{X}}(k+1|k)$ needs to be updated with the information provided by $\mathcal{Z}(k+1)$ in order to obtain the filtered estimate. When the measurements may be randomly delayed by one sampling time, the update equation for a linear filter using one step randomly delayed measurement is

$$
\hat{\mathcal{X}}(k+1|k+1) = \hat{\mathcal{X}}(k+1|k) + \bar{K}(k+1)(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k)), \tag{3.25}
$$

where $\bar{K}(k+1)$ is the filter gain and $\hat{\mathcal{Z}}(k+1|k)$ is a one step ahead prediction of $\mathcal{Z}(k+1)$ defined in (3.22). The estimation error covariance matrix of the state estimate $\hat{\mathcal{X}}(k+1|k+1)$ is given by

$$\bar{P}(k+1|k+1) = \mathbb{E}[\Phi(k+1)\Phi(k+1)^\top], \tag{3.26}$$

where $\Phi(k) := \mathcal{X}(k+1) - \hat{\mathcal{X}}(k+1|k+1)$.

The objective of this section is to find the filter gain $\bar{K}(k+1)$ in (3.25) that minimizes the trace of the estimation error covariance matrix $\bar{P}(k+1|k+1)$ of the state estimate $\hat{\mathcal{X}}(k+1|k+1)$. Our main result in this section is given in the next theorem.

Theorem 2. For system (3.23) with assumptions 1-2, the filter gain $\bar{K}(k+1)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(k+1|k+1)$ is given by

$$\begin{aligned}
&\bar{K}(k+1) = (\beta\bar{P}(k+1|k) + (1-\beta)A\bar{P}(k|k))C^\top[\beta(U_vU_v^\top + C\bar{P}(k+1|k)C^\top \\
&+ G_2q(k+1|k)G_2^\top) + (1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top + G_2q(k|k)G_2^\top) + \\
&\beta(1-\beta)(\widetilde{\psi}_0(k+1) + \widetilde{\psi}_1(k+1)) - \beta(1-\beta)(\psi_0(k+1)\psi_1(k+1)^\top \\
&+ \psi_1(k+1)\psi_0(k+1)^\top)]^{-1},
\end{aligned} \tag{3.27}$$

$$\begin{aligned}
\text{where}\quad &\bar{P}(k+1|k) = A\bar{P}(k|k)A^\top + U_wU_w^\top + G_1q(k|k)G_1^\top, \\
&\hat{\mathcal{Y}}(k+1-i|k) = C\hat{\mathcal{X}}(k+1-i|k) + D =: \psi_i(k+1), \\
&\widetilde{\psi}_i(k+1) = \psi_i(k+1)\psi_i(k+1)^\top, \quad i = 0, 1,
\end{aligned} \tag{3.28}$$

and $q(k|k)$ is as defined in equation (3.8) and $\bar{P}(k|k)$ is as defined in equation (A.13) in the Appendix.

Proof: See Appendix.

Remark 2. If $\beta = 1$, i.e., if there is no random delay, our result reduces to the filter derived in [56]. If, in addition to $\beta = 1$, we have $G_2 = 0$, i.e., if there is no multiplicative noise in the measurement equations, our result reduces to a special case of the filter derived in [77], with $\gamma = 1$ in the authors' notation in that paper. If, in addition to $\beta = 1$, we have $G_1 = G_2 = 0$, i.e., if there is no multiplicative noise, our filter reduces to the Kalman filter for the linear additive noise case.

Remark 3. We are seeking to minimize the variance over a set of filters where the additive correction term to the prediction is linear in the current measurement. We are not minimizing variance in the sense of finding $\mathbb{E}(\mathcal{X}(k)|\mathcal{Z}(k))$, which would mean having a very general (nonlinear) function of past and present values of $\mathcal{Z}$.

After describing the result for a single random delay case, we move on in the next section to a more general case where there might be up to $N > 1$ random delays at each time k, where N is an arbitrary but fixed integer.

Measurement with up to N random delays

We consider up to N consecutive delayed measurements. To account for these delays, we consider N i.i.d. Bernoulli random variables, p_k^i, $i = 1, 2, \cdots, N$, with possible values 0 or 1 as before. As in the previous section, it is assumed that $\mathcal{Y}(1)$ is available at time $k = 1$ for state estimation. At each time $k > 1$ the observation is either delayed by one or more sampling periods or is updated. Specifically, the measurement at time $k+1$ can be described as follows:

$$\begin{aligned}
\mathcal{Z}(k+1) &= (1 - p_{k+1}^1)\mathcal{Y}(k+1) + p_{k+1}^1(1 - p_{k+1}^2)\mathcal{Y}(k) + p_{k+1}^1 p_{k+1}^2(1 - p_{k+1}^3)\mathcal{Y}(k-1) \\
&+ \cdots + (\prod_{i=1}^{N-1} p_{k+1}^i)(1 - p_{k+1}^N)\mathcal{Y}(k+1-N+1) + (\prod_{i=0}^{N} p_{k+1}^i)\mathcal{Y}(k-N+1), \\
&= \sum_{i=0}^{N-1}(\prod_{j=0}^{i} p_{k+1}^j)(1 - p_{k+1}^{i+1})\mathcal{Y}(k+1-i) + (\prod_{i=0}^{N} p_{k+1}^i)\mathcal{Y}(k-N+1), \qquad (3.29)
\end{aligned}$$

with $p_{k+1}^0 := 1$. As a check, the same result is obtained for $N = 1$ as in the previous section. The proposed novel multiplicative structure of Bernoulli random variables ensures that only one measurement (current or delayed) is available at any time $k+1$, even though p_{k+1}^i are assumed to be independent of each other.

Remark 4. Using $\mathbb{E}(p_{k+1}^i) = 1 - \beta$ for $i = 1, 2, \ldots, N$ and using the fact that p_{k+1}^i are independent both from each other and from the measurement noise, we can easily obtain the following:

$$\hat{p}_k^i = \mathbb{E}\left(\prod_{j=0}^{i} p_{k+1}^j (1 - p_{k+1}^{i+1})\right)^2 = \beta(1-\beta)^i,$$

$$\hat{p}_k^N = \mathbb{E}\left(\prod_{i=0}^{N} p_{k+1}^i\right)^2 = (1-\beta)^N,$$

$$\widetilde{\hat{p}}_k^i = \mathbb{E}\left(\prod_{j=0}^{i} p_{k+1}^j (1 - p_{k+1}^{i+1}) - \beta(1-\beta)^i\right)^2 = \beta(1-\beta)^i - \beta^2(1-\beta)^{2i}, \tag{3.30}$$

$$\widetilde{\hat{p}}_k^N = \mathbb{E}\left(\prod_{i=0}^{N} p_{k+1}^i - (1-\beta)^N\right)^2 = (1-\beta)^N - (1-\beta)^{2N}. \tag{3.31}$$

Using Remark 4, we can write our one step ahead prediction of $\mathcal{Z}(k+1)$ in terms of the past predictions of $\hat{\mathcal{Y}}(k+1-i|k)$ as

$$\begin{aligned}\hat{\mathcal{Z}}(k+1|k) &= \beta\hat{\mathcal{Y}}(k+1|k) + (1-\beta)\beta\hat{\mathcal{Y}}(k|k) + (1-\beta)^2\beta\hat{\mathcal{Y}}(k-1|k) \\ &\quad + \cdots + (1-\beta)^{N-1}\beta\hat{\mathcal{Y}}(k-N+1|k) + (1-\beta)^N\hat{\mathcal{Y}}(k-N+1|k) \\ &= \beta\sum_{i=0}^{N-1}(1-\beta)^i\hat{\mathcal{Y}}(k+1-i|k) + (1-\beta)^N\hat{\mathcal{Y}}(k-N+1|k).\end{aligned} \tag{3.32}$$

where $\hat{\mathcal{Y}}(k+i|k)$ is defined by $\hat{\mathcal{Y}}(k+i|k) = C\hat{\mathcal{X}}(k+i|k)+D$. For the current set-up, the update equation for a linear filter using N steps randomly delayed measurement is

$$\hat{\mathcal{X}}(k+1|k+1) = \hat{\mathcal{X}}(k+1|k) + \bar{K}(k+1)(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k)), \quad (3.33)$$

where $\hat{\mathcal{X}}(k+1|k)$ is as defined in (3.24). Denoting the estimation error covariance matrix by $\bar{P}(k+1|k+1)$ as before, the expectation of terms of the form $\mathcal{X}(k-N)\mathcal{X}^\top(k)$ appears in $\bar{P}(k+1|k+1)$. To evaluate this expectation, the following result is useful.

Lemma 2. According to (3.1), the relationship between $\mathcal{X}(k)$ and $\mathcal{X}(k-N)$ is

$$\mathcal{X}(k-N) = A^{-N}(\mathcal{X}(k) - B - U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) + \sum_{i=1}^{N-1} A^{-(N-i)}(-B - U_w\mathcal{W}(k-(i+1)) - G_1 \operatorname{diag}(\mathcal{X}(k-(i+1)))\mathcal{S}_1(k-(i+1))),$$
$$N > 1. \quad (3.34)$$

Note: $\mathcal{X}(k-i)$, $i = 1, \ldots, N-1$, appearing on both sides of the equation here will not pose a problem in forming the expectation above since $\mathcal{S}_1(k)$ is a zero mean. i.i.d. sequence.

Before presenting the main result in this section, we first introduce the following useful lemma.

Lemma 3. The second order moment of $\operatorname{diag}(\mathcal{X}(k-i))$ is given as follows:

$$q(k-i|k) = \operatorname{diag}(vec(P_{jj}(k-i|k)) + (\hat{\mathcal{X}}(k-i|k))^2), \quad (3.35)$$

where $q(k-i|k) = \mathbb{E}[\operatorname{diag}(\mathcal{X}(k-i))\operatorname{diag}(\mathcal{X}(k-i))^\top]$,

$$
\begin{aligned}
&\hat{\mathcal{X}}(k-i|k) = \hat{\mathcal{X}}(k-i|k-1) + L(k-i)(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k-1)),\\
&\bar{P}(k-i|k) = \bar{P}(k-i|k-1) + L(k-i)\chi(k-1)L(k-i)-\\
&\zeta(k-i|k-1)L(k-i)^\top - L(k-i)\zeta(k-i|k-1)^\top,\\
&L(k-i) = \zeta(k-i|k-1)^\top \chi(k-1)^{-1},\\
&\zeta(k-i|k-1) = \sum_{i=0}^{N-1} \beta(1-\beta)^i (A^{-i}\bar{P}(k|k-1)(A^{-i+1})^\top - \sum_{j=0}^{i-1} A^{-(i-j)}(U_m U_m^\top +\\
&q(k-(j+1)|k-1)))(A^{-(i-j)+1}))^\top + (1-\beta)(A^{-i}\bar{P}(k|k-1)(A^{-N+1})^\top -\\
&\sum_{j=0}^{N-1} A^{-(N-j)}(U_m U_m^\top + q(k-(j+1)|k-1)))(A^{-(N-j)+1}))^\top),
\end{aligned} \tag{3.36}
$$

and $\chi(k-1)$ is as defined in equation (A.29) in the Appendix.

Proof: is on the same lines as Lemma 1.

We will derive an expression for $\bar{K}(k+1)$ which minimizes the trace of $\bar{P}(k+1|k+1)$. For this purpose, we assume that the i step ahead predictions $\hat{\mathcal{Y}}(k+1-i|k)$ obtained by using the MV filter are unbiased, i.e. $\mathbb{E}[\hat{\mathcal{Y}}(k+1-i|k)] = \mathcal{Y}(k+1-i)$. Note that this assumption is implicit in approximate nonlinear filters for systems with delays, such as [121]. Under this assumption, we have the following result:

Theorem 3. In the case when the measurement is randomly delayed by up to N time steps where $N > 1$, the filter gain $\bar{K}(k+1)$ that would minimize the trace of the estimation error covariance matrix $\bar{P}(k+1|k+1)$ is given by

$$\bar{K}(k+1) = (\beta\bar{P}(k+1|k)C^\top + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1}\beta(1-\beta)^i\times$$

$$(A\bar{P}(k|k)(A^{-i+1})^\top C^\top - \sum_{j=0}^{i-2}A^{j+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top)$$

$$+ (1-\beta)^N(A\bar{P}(k|k)(A^{-N+1})^\top C^\top - \sum_{i=0}^{N-2}A^{i+1}(U_wU_w^\top +$$

$$G_1q(k-(i+1)|k)G_1^\top)(A^{-(N-i)+1})^\top C^\top))\hat{\chi}(k)^{-1}, \tag{3.37}$$

where $\psi_i(k+1) = \hat{\mathcal{Y}}(k-i+1|k)$ as before, $q(k-i|k)$ is as defined in (3.35) and $\bar{P}(k+1|k)$, $\bar{P}(k|k)$ and $\widetilde{\psi}_i(k+1)$ are as defined in (A.28) in the Appendix. $\hat{\chi}(k)$ is as defined in equation (A.29) in the Appendix.

Proof: See Appendix.

Remark 5. Note that it is conceptually easy to generalize this algorithm to a situation where the probabilities of delays for different sample times differ from one another (see, e.g., [122] for an example with two delays). We do not do this for two reasons. Firstly, it complicates the notation even further and formulae considerably without adding much value. Secondly and more importantly, the parameters representing the delay probabilities are not easy to identify from data. Identifying a single parameter as a rate of arrival (e.g., $1-\beta$ missing measurements per 100 sample times) and then using it as a proxy rate for a Poisson process truncated at N possible delays seems to be a sensible compromise. In previous research dealing with delays and multiplicative noise in [112] and [123], note that multiplicative noise is modelled as a norm bounded uncertainty, which is a fundamentally different approach from our assumption of treating it as stochastic noise.

In the next section, we demonstrate the application of this algorithm for both a single random delay and for multiple random delays.

3.3.3 Numerical examples

Example 1

Consider a linear state space system with additive-multiplicative noise given by (3.1)-(3.2), with $A = -0.5$, $B = 0$, $C = 0.45$, $D = 0$, $U_w = 0.1$, $U_v = 0.6$, $G_1 = 0.1$ and $G_2 = 0$. As in section 4.1.3, $\mathcal{W}(k)$, $\mathcal{V}(k)$, $\mathcal{S}_1(k)$ and $\mathcal{S}_2(k)$ are uncorrelated random variables with $\mathbb{E}(\mathcal{V}(k)) = \mathbb{E}(\mathcal{W}(k)) = \mathbb{E}(\mathcal{S}_1(k)) = \mathbb{E}(\mathcal{S}_2(k)) = 0$ and $\mathbb{E}(\mathcal{V}(k))^2 = \mathbb{E}(\mathcal{W}(k))^2 = \mathbb{E}(\mathcal{S}_1(k))^2 = \mathbb{E}(\mathcal{S}_2(k))^2 = 1$. In the simulation, we consider two different sets of the initial conditions. The first set of initial conditions are $\mathcal{X}(0) = 1$, $\hat{\mathcal{X}}(0) = 0$ and $P(0) = 1$ (Case I). The second set of initial conditions are $\mathcal{X}(0) = 0.1$, $\hat{\mathcal{X}}(0) = 0.65$ and $P(0) = 10$ (Case II).

We perform two experiments for this system with different values of β. Firstly, we consider a situation where the measurement might be delayed by up to two sample times, that is

$$\mathcal{Z}(k) = (1-p_k^1)\mathcal{Y}(k)+p_k^1(1-p_k^2)\mathcal{Y}(k-1)+p_k^1 p_k^2\mathcal{Y}(k-2) \quad k > 1, \quad \mathcal{Y}(1) = \mathcal{Z}(1).$$

In order to compare the performance of the estimators, we use the RMSE criteria. Consider 100 independent simulations, each with 200 data points. Denoting $\mathcal{X}^{(s)}(k), \quad k = 1, ..., 200$ as the s^{th} set of true values of the state and $\hat{\mathcal{X}}^{(s)}(k|k)$ as the filtered state estimate at time k for the s^{th} simulation run, the RMSE of the filter for each of the algorithms is calculated by

$$RMSE(s) = \sqrt{\frac{1}{200}\sum_{k=1}^{200}(\mathcal{X}^{(s)}(k) - \hat{\mathcal{X}}^{(s)}(k|k))^2}, \qquad s = 1, ..., 100.$$

Then, the average of the RMSE for each of the algorithms over 100 simulations is given by

$$AvRMSE = \frac{1}{100}\sum_{s=1}^{100} RMSE(s).$$

The data to test filtering performance was generated by simulation with two random delays. We compare the performance of three filters for this data generating system with two delays: the filter in [77] (i.e., there is no random delay), a filter designed for a single delay and a filter for two delays. The results of the MV filter with the filter in [77], one delay and two delays are represented by DKF_0, DKF_1 and DKF_2, respectively. The rationale behind comparing these filters is as follows:

1. To my knowledge, my proposed algorithm is the only systematic way of dealing with random delays under additive-multiplicative noise, so that it makes sense to compare our proposed algorithm with a delay-free filter (see [77]) with the same noise description. This justifies the comparison between DKF_0 and DKF_2.

2. Further, as we have commented elsewhere, it is not easy to estimate accurately the maximum number of random delays in the system. It is therefore of interest to see how the filter performs with an 'incorrect' number of delays assumed; in particular, it is of interest to see if DKF_1 performs any better than DKF_0 when there is a random delay in the system even if the maximum number of delays exceeds 1.

The AvRMSE values are calculated for these filters with different values of β. The results are summarized in Table 3.4 for Case I and in Table 3.5 for Case II. We can see from these tables that, in all the cases, the estimators which account for delays perform better than the estimator with no delay proposed in [77]. In particular, the errors for the filter designed for the correct maximum number of random delays (i.e., DKF_2) are the lowest of all the cases, followed by the errors for DKF_1 and the errors for DKF_0. The reason is that the filter in [77] does not use the knowledge of delays. The fact that DKF_1 gives lower errors that DKF_0 is non-trivial and indicates that, in the presence of random delays, even using a possibly incorrect maximum number of random delays (e.g., assuming a single random delay) may still be better than ignoring random delays altogether.

Table 3.4: Comparison of the $AvRMSE$ for different values of β Case I (Example 1)

		$\beta = 0.9$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.3$
$AvRMSE$	DKF_0	0.1152	0.1153	0.1160	0.1175
	DKF_1	0.1149	0.1150	0.1155	0.1167
	DKF_2	0.1141	0.1145	0.1149	0.1157

Table 3.5: Comparison of the $AvRMSE$ for different values of β Case II (Example 1)

		$\beta = 0.9$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.3$
$AvRMSE$	DKF_0	0.1156	0.1168	0.1172	0.1188
	DKF_1	0.1154	0.1159	0.1168	0.1176
	DKF_2	0.1149	0.1153	0.1164	0.1171

In the second experiment, it is assumed that the observations are randomly delayed by one sampling period:

$$\mathcal{Z}(k) = (1 - p_k)\mathcal{Y}(k) + p_k\mathcal{Y}(k-1), \quad k > 1, \quad \mathcal{Y}(1) = \mathcal{Z}(1).$$

We compare the performance of the proposed algorithm and the algorithm proposed in [109]. The AvRMSE is calculated for these filters with different values of β for Case I. The results are summarized in Table 3.6. We can see from this table that, in all the cases, the proposed algorithm perform better than the algorithm proposed in [109]. From two experiments, we note that the AvRMSE becomes smaller as the probability of the on-time arrival becoming larger which is in keeping with the comments in [119] and [118].

Example 2

Consider again the discrete time system in example 2 of section 3.2.3, which is repeated here for ease of reference:

Table 3.6: Comparison of $AvRMSE$ for our filter and filter in [109] for different values of β Case I (Example 1)

		$\beta = 0.9$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.3$
$AvRMSE$	filter in [109]	0.1195	0.1211	0.1225	0.1245
	our filter	0.1180	0.1192	0.1201	0.1231

$A = \rho \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ \sin(\lambda) & \cos(\lambda) \end{bmatrix}, \quad B = 0, \quad C = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad D = 0,$
$U_w^2 = 214$, $U_v^2 = 1593$ and $G_2 = 0$. ρ and λ are random variables with means $\hat{\rho} = 0.4$ and $\hat{\lambda} = 0.41$. The dynamics and the chosen parameter values (with $G_1 = G_2 = 0$) are used in ([86], chapter 2) as a time series model of rainfall in north-east Brazil. As the parameters are obtained from data, one usually has an estimate of the standard error in each parameter. We allocate standard errors to λ and ρ (as 10% of nominal values) and compute G_1 via Monte Carlo simulation. This is obtained as

$$G_1 = \begin{bmatrix} 0.2168 & 0.0001 \\ 0.0001 & 0.6360 \end{bmatrix}.$$

The initial conditions are
$\mathcal{X}(0) = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top, \quad \hat{\mathcal{X}}(0) = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$ and $P(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$
The AvRMSE for $\mathcal{Y}(k) = \mathcal{X}_1^s(k) + \mathcal{X}_2^s(k)$ in this case is calculated as follows:

$$RMSE(s) = \sqrt{\frac{1}{200} \sum_{k=1}^{200} ((\mathcal{X}_1^{(s)}(k) + \mathcal{X}_2^{(s)}(k)) - (\hat{\mathcal{X}}_1^{(s)}(k|k) + \hat{\mathcal{X}}_2^{(s)}(k|k))^2}, \qquad s = 1, ..., 100.$$

Then the average of RMSE for $\mathcal{Y}(k) = \mathcal{X}_1^s(k) + \mathcal{X}_2^s(k)$ over 100 simulations is given by

$$AvRMSE = \frac{1}{100} \sum_{s=1}^{100} RMSE(s).$$

As before, we perform two experiments for this system with different values of β. Firstly, we compare the performance of three filters for this data generating system with two delays: a no delay filter proposed in [77], a filter designed for a single delay and a filter for two delays. Even though this delay scenario is not the most realistic, our purpose is to illustrate that the new filtering algorithm is robust to delays. Table 3.7 provide the errors for various values of β with the filter in [77] which accounts for multiplicative noise, one delay filter and a two delay filter which clearly illustrate the superior performance of our algorithm. As in the previous example, the results of the MV filter with the filter in [77], one delay and two delays are represented by DKF_0, DKF_1 and DKF_2, respectively. As can be seen, the filter with two delays outperforms the other filters and the filter with a single delay still outperforms the filter with no delays proposed in [77] for all values of β and for both measures of error.

Table 3.7: Comparison of $AvRMSE$ for DKF_0, DKF_1 and DKF_2 for different values of β (Example 2)

		$\beta = 0.9$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.3$
$AvRMSE$	DKF_0	28.1939	29.0632	29.1362	29.7086
	DKF_1	27.5920	28.4413	28.4509	28.9511
	DKF_2	27.3136	28.2516	28.4135	28.7595

In the second experiment, we compare the performance of DKF_1 and DKF_2 with the algorithm proposed in [110]. The AvRMSE is calculated for these filters with different values of β. The results are summarized in Table 3.8. We can see from this table that, in all the cases, the proposed algorithm (i.e., DKF_1 and DKF_2) perform better than the algorithm proposed in [110].

Table 3.8: Comparison of $AvRMSE$ for DKF_1 and DKF_2 with filter in [110] for different values of β (Example 2)

		$\beta = 0.9$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.3$
$AvRMSE$ $AvMRAE$	Filter in [110]	27.5880	27.5910	28.3658	28.7343
	DKF_1	27.2414	27.4738	28.1173	28.4261
	Filter in [110]	27.0773	27.4602	28.2203	28.5127
	DKF_2	27.0695	27.4563	28.1056	28.3414

3.4 Summary

In this chapter a class of discrete time systems with both additive and multiplicative noise is considered. The optimal MV filter which is linear in the current measurement is discussed for this class of systems. The closed-form solution generalizes the results for MV filtering for additive-multiplicative noise case in [77]. We have also provided an interpretation of filtering under multiplicative noise in terms of filtering under parameter perturbations in an additive noise model. A new filter has also been considered for state estimation problems in a class of discrete time systems with both additive and multiplicative noise when the measurement might be delayed by one or more sampling times. The filter minimizes the trace of state covariance matrix over a class of filter gain matrices which generalize the Kalman filter gain in a specific sense. The optimization is carried out on a set of filters which have a Kalman filter-like structure, although the gain matrix itself depends on the previous values of (possibly delayed) measurements. The proposed closed-form solution generalizes the results for MV filtering for the additive-multiplicative noise case in [56] (which does not treat delays). The results of this chapter were applied to four different real data experiments for linear systems with additive-multiplicative noises. The first two numerical examples illustrate the utility of the proposed algorithm with parameter perturbations. Our numerical experiments indicate that the proposed filtering algorithm can be used to improve filtering performance as measured by the estimation error variance when there is uncertainty in the estimated model parameters. The third and fourth numerical experiments illustrate the utility of the proposed algorithm with additive-multiplicative noises, where the measurement might be delayed randomly by one and two sample times. Our numerical experiments indicate that the proposed algorithm outperforms an implementation which ignores delays for a range of delay probabilities. The proposed filter can be applied to state estimation of a system in which the estimator consists of sensors connected through communication networks and many real-world applications where measurements are delayed randomly and there is parametric uncertainty to be accounted for. We have also demonstrated that using a filter with a single random delay might still be useful and better than ignoring random delays if the maximum number of random delays is unknown. Coupled with a local linearisation of any fully nonlinear system, this algorithm has a potential to be a very useful tool in signal estimation across a wide range of fields where delays are an issue.

Chapter 4

Approximate minimum variance filter under random delays II: linear continuous discrete systems with additive and multiplicative noise

4.1 Introduction

In chapter 4 the problem of the latent state estimation for linear discrete time systems with randomly delayed observations and additive-multiplicative noise was studied. In this chapter the problem will be considered for continuous discrete systems. The material presented in this chapter has been published in [124] and submitted for publication in [125].

4.2 A minimum variance filter for continuous discrete systems with additive-multiplicative noise

4.2.1 Background

The Bayesian filtering framework is a robust filtering framework where the state dynamics are typically modelled with stochastic differential or difference equations. In *continuous-discrete filtering* problems (see, e.g., [126] and [127]) the process noise is represented in continuous time and the measurement equation is given in discrete time, *i.e.* the measurements (which are typically noisy) are available at discrete time instants. The measurement model in the continuous discrete filter is thus of the same form as in discrete-time filtering. The measurement frequency may be limited by hardware or other physical considerations. The major difference between the continuous discrete filter (henceforth abbreviated as CDF) and the discrete time filter (or discrete discrete filter, abbreviated as DDF) is that, in the DDF approach, both the state dynamics and the noisy measurements are modelled as discrete-time processes. Filtering problems where a continuous-time signal is observed discretely in time have received a great deal of attention, since this kind of formulation often arises in numerous applications such as GPS and inertial navigation [128], stochastic control [129], target tracking [130] and finance [131]. The Bayesian optimal CDF (e.g., [126] and [132]) is the same as the DDF only when measurements are obtained at discrete time instants, the posterior density is propagated from one sampling instant to the next by solving the associated Fokker-Planck equation. In the literature, many conventional filtering algorithms are extended to deal with continuous-discrete systems. Examples of such algorithms include EKF [126], which approximates the exact solution by using a Taylor series expansion approximation to the nonlinear drift function and forms a Gaussian process approximation to the SDE, PF [133], where a set of weighted particles is used for approximating the posterior probability measure and UKF [134], which relies on propagating a set of points representing a Gaussian density with the correct first two moments through the system equations. In [135], the extension of the cubature Kalman filter [136] to continuous-discrete filtering using Itô-Taylor expansion of continuous dynamics is studied. The results in [137], which use the cubature integration method in continuous-discrete filtering, were generalized in [138]. In [139], closed-form solutions of continuous-discrete systems are derived. In [140], CDF algorithms using the EKF, UKF and PF with applications to the angle-only tracking in 3D are developed. Most of the results mentioned above are concerned with additive noise only and multiplicative noise (or any other state-dependent noises) are not taken into account. For systems subject to multiplicative noise, different kinds of algorithms have been introduced for continuous-discrete time models. These algorithms are reported in [141], [142] and [143]. The optimal linear one-stage prediction, filtering and smoothing in the case of continuous-time system are derived in [144], where the observation matrix is multiplied by scalar binary-valued white noise. In [145], the filtering problem for continuous-discrete linear state space models is considered, where the solution is given in form of solving coupled ODEs. In [146], both deterministic and stochastic perturbations are considered in the design of an H_∞-type theory for continuous-time stochastic linear plants. The problem of Kalman filtering for a class of uncertain linear continuous-time systems with Markovian jumping parameters is studied in [147], where the system is subjected to time-varying norm-bounded parameter uncertainties in the state and measurement equations.

The motivation of this section is to extend the result in section 3.1 to deal with continuous discrete problems. Specifically, we consider a class of continuous discrete systems with both additive and multiplicative noise, including square-root affine systems. In this section, we use the Euler scheme followed by conditional moment matching to transform SDEs in the process equation into a discrete model on a timescale which is finer than the measurement timescale. The problem addressed here is the design of a filter that minimizes the trace of the estimation error covariance matrix at each measurement sampling instant as well as at points in-between the measurement sampling instants. We demonstrate through numerical experiments that our new filter performs better than the corresponding DDF in [77] when information about continuous time dynamics is available.

4.2.2 System model and problem formulation

State space model

Consider a system in which the process equation is described by a stochastic differential equation:

$$d\mathcal{X}(t) = (A\mathcal{X}(t) + B)dt + U_w d\mathcal{W}(t) + G\,\mathrm{diag}(\mathcal{X}^{\gamma}(t))d\mathcal{S}(t) \tag{4.1}$$

The behaviour of the system is observed through noisy measurements $\mathcal{Y}(t_k)$ which are taken at discrete time instants $t_k = kT$, where T is the measurement sampling interval:

$$\mathcal{Y}(t_k) = C\mathcal{X}(t_k) + U_v\mathcal{V}(t_k). \tag{4.2}$$

Here, $\gamma \in \{0, 0.5, 1\}$, $\mathcal{X}(t)$ is an n-dimensional state of the system at any time t, $\mathcal{Y}(t_k) \in \mathbb{R}^r$ is the measurement at the t_k^{th} time instant and A, B, G, C, U_w and U_v are given constant matrices of appropriate dimensions. $\mathcal{X}^{\gamma}(t)$ indicates a vector whose each element is the corresponding element of $\mathcal{X}(t)$ raised to the power γ. $\mathcal{W}(t) \in \mathbb{R}^n$ is a standard Wiener process with increment $d\mathcal{W}(t)$, and $\mathcal{V}(t_k), k = 1, 2, \cdots$ is a discrete time stochastic process which represents the measurement noise. The standard Wiener process $\mathcal{S}(t) \in \mathbb{R}^n$ represents the multiplicative noise. The initial state is a random vector with a known mean and covariance matrix, $\mathbb{E}[\mathcal{X}(0)] = \hat{\mathcal{X}}(0)$ and $\mathbb{E}[(\mathcal{X}(0)-\hat{\mathcal{X}}(0))(\mathcal{X}(0)-\hat{\mathcal{X}}(0))^{\top}] = P(0)$, respectively. $\mathcal{X}(0)$, $\mathcal{W}(t)$, $\mathcal{V}(t_k)$ and $\mathcal{S}(t)$ are mutually independent. This class of systems includes systems with additive noise ($\gamma = 0$), multiplicative noise ($\gamma = 1$) and square root affine noise ($\gamma = 0.5$). The last case leads to noise terms for which covariance is affine in the state variables, and is especially relevant in financial mathematics; see, e.g., [148] and references therein.

The purpose of optimal (Bayesian) CDF is to determine the evolution in time t of the conditional PDF , also called the posterior density of the state defined for all $t \geq 0$:

$$p(\mathcal{X}(t)|\mathcal{Y}(t_1), ..., \mathcal{Y}(t_{k+1})), \quad t \in [t_k, t_{k+1}], \quad k = 1, 2, \dots,$$

or at least the relevant moments of the distribution (e.g., the mean vector and the covariance matrix). The optimal continuous-discrete Bayesian filter is the same as the DDF performed in two steps.

1) Prediction step: In this step, the prior PDF is evaluated by propagation of the previous posterior density between the measurement instants.

2) Update step: In this step, the posterior density is obtained by updating the predictive density using the measurement and Bayesian rule. This step is the same as the DDF update step because the measurement update relies only on the measurement equation, which is modelled in discrete time for a continuous-discrete state-space model case.

Solving the dynamic system (4.1)-(4.2) is very challenging since the SDEs appearing in the dynamic model or the corresponding Fokker-Planck-Kolmogorov partial differential equations cannot typically be solved analytically, and approximation must be used. We look at reducing the state transition equation to a discrete form at a higher sampling frequency than the measurement frequency (or a smaller time step size than the measurement sampling step size) and then adapting existing techniques to deal with the resulting system.

Discretisation of process model

Let $t_{k+1} - t_k = \delta$, $k > 0$ be the uniform time interval between consecutive measurement samples. Applying the Euler scheme to (4.1) over time interval

$(t_k, t_k + \Delta)$ yields

$$\mathcal{X}(t_k^{i+1}) = \mathcal{X}(t_k^i) + (A\mathcal{X}(t_k^i) + B)\Delta + U_w \Delta W + G\,\mathrm{diag}(\mathcal{X}^\gamma(t_k^i))\Delta S,$$

where $t_k^i \in [t_k, t_{k+1}], i = 0, 1, 2, \ldots, m-1$ are uniformly spaced inter-sampling times, $t_k^m = t_{k+1}$ and $\Delta = \frac{\delta}{m}$. ΔW and ΔS are n-dimensional Gaussian random variables with zero mean and covariance matrices $\mathbb{E}[\Delta W \Delta W^\top] = \Delta\mathcal{I}$, $\mathbb{E}[\Delta S \Delta S^\top] = \Delta\mathcal{I}$, respectively.

In order to match exactly the first two moments, we use the moment matching approach. So, the expression for the conditional mean of $\mathcal{X}(t_k^{i+1})$ given $\mathcal{X}(t_k^i)$ can be easily shown to be

$$\mathbb{E}[\mathcal{X}(t_k^{i+1})|\mathcal{X}(t_k^i)] = \mathcal{X}(t_k^i) + (A\mathcal{X}(t_k^i) + B))\Delta,$$

where $i = 0, 1, \ldots, m-1$ and the associated conditional covariance matrix is

$$var[\mathcal{X}(t_k^{i+1})|\mathcal{X}(t_k^i)] = U_w U_w^\top \Delta + G(\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i)))^2 G^\top \Delta.$$

Then,

$$\mathcal{X}(t_k^{i+1}) = \tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m \tilde{\mathcal{W}}(t_k^i) + \tilde{G}\,\mathrm{diag}(\mathcal{X}^\gamma(t_k^i))\tilde{\mathcal{S}}(t_k^i) \tag{4.3}$$

where

$$\tilde{A} = I + A\Delta,\ \tilde{B} = B\Delta,\ \tilde{U}_m = U_w\sqrt{\Delta},\ \tilde{G} = G\sqrt{\Delta},$$

and $\tilde{\mathcal{W}}$ and $\tilde{\mathcal{S}}$ are uncorrelated, zero mean random processes with identity covariance matrices. This puts the system in a discrete state space framework to which the standard discrete time filtering tools can be applied.

Minimum variance filter for the continuous-discrete system

To derive the recursive filtering equations, it is assumed that the observations are given up to time t_k and that the conditional mean of $\mathcal{X}(t_k^i)$ given $\mathcal{Y}(t_k)$, $\hat{\mathcal{X}}(t_k^i|t_k)$, is available. From this value, the conditional mean of $\mathcal{X}(t_k^{i+1})$, which provides the predictor $\hat{\mathcal{X}}(t_k^{i+1}|t_k)$ is derived using (4.3):

$$\hat{\mathcal{X}}(t_k^{i+1}|t_k) = \tilde{A}\hat{\mathcal{X}}(t_k^i|t_k) + \tilde{B}. \tag{4.4}$$

The predicted estimate $\hat{\mathcal{X}}(t_k^{i+1}|t_k)$ needs to be updated with the information provided by $\mathcal{Y}(t_{k+1})$ to obtain the filtered estimate. So, the update equation for a linear filter is

$$\hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}) = \hat{\mathcal{X}}(t_k^{i+1}|t_k) + \bar{K}(t_{k+1}^i)(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k)), \tag{4.5}$$

where $\hat{\mathcal{X}}(t_k^{i+1}|t_{k+1})$ indicates updated estimate of $\hat{\mathcal{X}}(t_k^{i+1})$ after $\mathcal{Y}(t_{k+1})$ becomes available, and the estimation error covariance matrix is given by

$$\bar{P}(t_k^{i+1}|t_{k+1}) = \mathbb{E}[(\mathcal{X}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}))((\mathcal{X}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}))^\top]. \tag{4.6}$$

As in chapter 3, our objective is to find a filter gain $\bar{K}(t_{k+1}^i)$ that would minimize the trace of the estimation error covariance matrix $\bar{P}(t_k^{i+1}|t_{k+1})$ of the state estimate $\hat{\mathcal{X}}(t_k^{i+1}|t_{k+1})$ and obtain an expression for the filter. Our main result in this section, which is an extension of the corresponding result from [77], is given in the next theorem.

Theorem 4. For system (4.2) (4.3), the filter gain $\bar{K}(t_{k+1}^i)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(t_k^{i+1}|t_{k+1})$ is given by

$$\begin{aligned}\bar{K}(t_{k+1}^i) &= (\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k) - \tilde{U}_m\tilde{U}_m^\top - \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}^\gamma(t_k^{m-1}))\operatorname{diag}(\mathcal{X}^\gamma(t_k^{m-1}))^\top]\times \\ &\tilde{G}^\top)C^\top - \sum_{r=1}^{m-i-2}\tilde{A}^{-(m-i-r)+1}(\tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}^\gamma(t_k^{m-(r+1)}))\operatorname{diag}(\mathcal{X}^\gamma(t_k^{m-(r+1)}))^\top]\times \\ &\tilde{G}^\top)(\tilde{A}^{(r)})^\top C^\top)[C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top]^{-1},\end{aligned} \tag{4.7}$$

where

$$\bar{P}(t_k^m|t_k) = \bar{P}(t_{k+1}|t_k) = \tilde{A}\bar{P}(t_k|t_k)\tilde{A}^\top + \tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}^\gamma(t_k))\operatorname{diag}(\mathcal{X}^\gamma(t_k))^\top]\tilde{G}^\top \tag{4.8}$$

and

$$\begin{aligned}
&\mathbb{E}[\,\mathrm{diag}(\mathcal{X}^{\gamma}(t_k))\,\mathrm{diag}(\mathcal{X}^{\gamma}(t_k))^{\top}] \\
&= \mathrm{diag}\,(vec(P_{jj}(t_k|t_k)) + (\hat{\mathcal{X}}(t_k|t_k))^2) \qquad \text{if} \qquad \gamma = 1, \\
&= \mathrm{diag}\,(\hat{\mathcal{X}}(t_k|t_k)) \qquad \text{if} \qquad \gamma = 0.5, \\
&= \mathrm{diag}\,(I_n) \qquad \text{if} \qquad \gamma = 0.
\end{aligned} \tag{4.9}$$

Proof: See Appendix.
A few remarks on this result are in order.

- If $G = 0$ i.e., if there is no multiplicative noise, our filter reduces to the Kalman Bucy filter for continuous-discrete models with additive noise [1].

- The major difference between my work and the work represented in [77] is that in this section the dynamics are modelled as a continuous-time process and the measurements are modelled as a discrete-time process, while in [77] both the dynamics and measurements are modelled as discrete-time processes. This requires updating the values of $\mathcal{X}(t)$ at $t \in (t_k, t_{k+1})$. As the numerical examples show, this improves the prediction quality. If $m = 1$, we recover the results from [77].

- As mentioned earlier, the three specific values of γ, viz. $0, 0.5$ and 1, encompass the cases of additive, square root affine and multiplicative noise, respectively. As seen in theorem 4 (see equation (4.9), in particular), these choices of γ still allow us to obtain a closed-form recursive expression for the covariance matrix. Further, note that $\gamma \leq 1$ in our set-up is sufficient for the Euler scheme to converge; see [60], chapter 10, for example.

4.2.3 Numerical example

Example 1

For numerical evaluation, we used the same model parameters as described in example 1 of section 3.3.3, which is repeated here for ease of reference: $A = -0.5$, $B = 0$, $C = 0.45$, $U_w = 0.1$, $U_v = 0.6$ and $G = 0.1$. $\mathcal{W}(t)$ and $\mathcal{S}(t)$ are standard Wiener processes, $\mathcal{V}(t_k)$ is a random variable with $\mathbb{E}(\mathcal{V}(t_k)) = 0$ and $\mathbb{E}(\mathcal{V}(t_k))^2 = 1$ and is uncorrelated with $\mathcal{W}(t)$ and $\mathcal{S}(t)$. the initial conditions are $\mathcal{X}(0) = 1$, $\hat{\mathcal{X}}(0) = 0$ and $P(0) = 1$.

The measurement sampling period is $\delta = 1$. We consider a sequence of two different time steps between $t_k = k\delta$ and $t_{k+1} = (k+1)\delta$, $m = 5$ and $m = 10$, so $\Delta = 1/5$ and $\Delta = 1/10$. We then use these parameters to derive the discretisation parameters presented in section 5.2.2. That is,

$$\tilde{A} = 0.9, \quad \tilde{B} = 0, \quad \tilde{U}_w = 2.6833, \quad \tilde{G} = 0.0447,$$
$$\tilde{A} = 0.95, \quad \tilde{B} = 0, \quad \tilde{U}_w = 1.8974, \quad \tilde{G} = 0.0316,$$

for $\Delta = 1/5$ and $\Delta = 1/10$, respectively. In order to compare the performance of the estimators, we use the RMSE criteria. Consider 100 independent simulations, each with 200 data points. Denoting $\mathcal{X}^{(s)}(t_k), \quad k = 1, ..., 200$ as the s^{th} set of true values of the state and $\hat{\mathcal{X}}^{(s)}(t_k|t_k)$ as the as the filtered state estimate at time t_k for the s^{th} simulation run, the RMSE of the filter for each of the algorithms is calculated by

$$RMSE(s) = \sqrt{\frac{1}{200}\sum_{t_k=1}^{200}(\mathcal{X}^{(s)}(t_k) - \hat{\mathcal{X}}^{(s)}(t_k|t_k))^2}, \quad s = 1, ..., 100.$$

Then, the average of RMSE for the state over 100 simulations is given by

$$AvRMSE = \frac{1}{100}\sum_{s=1}^{100} RMSE(s).$$

Using our estimator, we compare the performance of two filters: the CDF and the DDF (*i.e.,* filter presented in [77]) with different values of γ. The results of the continuous-discrete filter and the discrete discrete filter are represented in the tables by CDF and DDF, respectively. The results are summarized in Tables 4.1 and 4.2. We can see from these tables that, in all cases, the estimators with CDF perform better than the estimator with DDF.

Table 4.1: Comparison of $AvRMSE$ for CDF and DDF for different values of γ and with $\Delta = 1/5$

		$\gamma = 0$	$\gamma = 0.5$	$\gamma = 1$
$AvRMSE$	CDF	2.1682	2.1189	2.0991
	DDF	4.1188	3.8428	3.7547

Table 4.2: Comparison of $AvRMSE$ for CDF and DDF for different values of γ and with $\Delta = 1/10$

		$\gamma = 0$	$\gamma = 0.5$	$\gamma = 1$
$AvRMSE$	CDF	1.5691	1.5349	1.5316
	DDF	3.5794	3.7215	3.6086

Example 2

As another example, we consider the same model parameters as described in [77] for the system (4.1)-(4.2) which are:

$$A = \begin{bmatrix} 0 & -0.5 \\ 1 & 1 \end{bmatrix}, \quad B = 0 \quad U_w = \begin{bmatrix} -6 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} -100 & 10 \end{bmatrix},$$

$$G = \begin{bmatrix} 0.12 & 0.02 \\ 0.15 & 0.1 \end{bmatrix},$$

and $U_v = 1$. As before, $\mathcal{W}(t)$ and $\mathcal{S}(t)$ are standard Wiener processes,

$\mathcal{V}(t_k)$ is random variable with zero mean and identity covariance matrix $\mathcal{I}$ and uncorrelated with $\mathcal{W}(t)$ and $\mathcal{S}(t)$. The initial conditions are
$\mathcal{X}(0) = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$, $\quad \hat{\mathcal{X}}(0) = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$ and $P(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
The measurement sampling period is $\delta = 1$. We consider a sequence of 10 time steps between $t_k = \delta k$ and $t_{k+1} = (k+1)\delta$, so that $\Delta = 1/10$. A difference in our simulation was that instead of using the parameters described in [77] themselves, we used them to derive the discretisation parameters presented in section 4.2.2. That is,

$$\tilde{A} = \begin{bmatrix} 0.9 & -0.05 \\ 0.1 & 0.9 \end{bmatrix}, \quad \tilde{B} = 0 \quad \tilde{U}_w = \begin{bmatrix} -1.8974 \\ 0.3162 \end{bmatrix}, \quad \tilde{G} = \begin{bmatrix} 0.0379 & 0.0063 \\ 0.0474 & 0.0316 \end{bmatrix},$$

We compare the performance of two filters for this data generating system: the CDF and the DDF (*i.e.*, filter presented in [77]) with different values of γ. In keeping with the notation in the previous example, the results of the continuous-discrete filter and discrete discrete filter will be represented by CDF and DDF, respectively. The AvRMSE is calculated for these filters with different values of γ, as in the previous subsection. Table 4.3 summarizes the results of this experiment. As can be seen, the CDF provides better accuracy when compared to DDF in all cases.

Table 4.3: Comparison of $AvRMSE_1$ and $AvRMSE_2$ for CDF and DDF for different values of γ

		$\gamma = 0$	$\gamma = 0.5$	$\gamma = 1$
$AvRMSE_1$	CDF	2.1008	2.0561	2.1588
	DDF	2.8492	2.8364	3.0235
$AvRMSE_2$	CDF	0.3037	0.3027	0.3703
	DDF	0.3729	0.3764	0.3945

4.3 A minimum variance filter for continuous-discrete linear systems with randomly delayed observations and additive and multiplicative noise

4.3.1 Background

All the results reported in the previous section depend on the assumption that the measurement signals are perfectly transmitted. However, in various practical situations the measurement available in the estimation may be not up to date due to the several factors like slow sensors, long processing time of the sensor data, limited capacity of the communication link, *etc.* In chapter 3, the delay is considered to be an integer multiple of the sampling time and a fractional delay is not taken into account. Further, it is noted that most available results with respect to filtering problems with randomly delayed observations have centered on discrete-time systems, and research into continuous systems for these problems have been very few. In recent years, however, the estimation of the state of continuous-discrete systems with randomly delayed observations has received considerable research attention and many important results have been reported. In a very recent study [55], the continuous-discrete filtering problem for randomly delayed measurements is investigated. By using reorganized innovation analysis, [149] designed the minimum mean square error (MMSE) estimation problem for linear continuous time varying systems with time delay measurements. In [150], the mean square stochastic stability for continuous time systems with stochastic delays have been investigated. Robust H_∞ filtering of uncertain systems with state delays has been considered in [151].

The literature on filtering problems for continuous-time systems from delayed observations as well as multiplicative noise is far less extensive. In [152], linear filtering for continuous-time systems with time-delayed measurements and multiplicative noises is presented using reorganized innovation. This method is based on solving two Riccati equations in real time for each estimation step. The problem of robust H_∞ filtering for Markovian jump linear systems with parameter uncertainties and mode-dependent time-varying delays has been studied in [153], where a linear matrix inequality (LMI) approach has been developed to design a Markovian jump linear filter. In [154], the problem of robust H_∞ filtering for continuous-time uncertain linear systems with multiple time-varying delays in the state variables is investigated.

Up to now, to the best of my knowledge, MV filtering for continuous discrete systems in the presence of both additive and multiplicative noise when the measurements are randomly delayed has not yet been studied, which motivates the present study. In particular, my work extends the results in [124] to deal with delayed measurements, using the theory developed for discrete linear systems in chapter 3. While in the previous chapter a filter with a Kalman filter-like structure is designed to approximately minimize the trace of the estimation error covariance matrix at t_{k+1}, whereas the filter in this chapter seeks to minimize trace of the estimation error covariance matrix at points in-between the measurement sampling instants as well as at t_{k+1}.

4.3.2 System model and problem formulation

State space model

The process equation can be described by the stochastic differential equation

$$d\mathcal{X}(t) = (A\mathcal{X}(t) + B)dt + U_w d\mathcal{W}(t) + G_1 \operatorname{diag}(\mathcal{X}(t))d\mathcal{S}_1(t). \tag{4.10}$$

The behaviour of the system is observed through the noisy measurements $\mathcal{Y}(t_k)$ which are taken at the discrete time instants $t_k = kT$, $k = 1, 2, \cdots$, where T is the measurement sampling interval:

$$\mathcal{Y}(t_k) = C\mathcal{X}(t_k) + U_v\mathcal{V}(t_k) + G_2 \operatorname{diag}(\mathcal{X}(t_k))\mathcal{S}_2(t_k), \tag{4.11}$$

where $\mathcal{X}(t)$ is the n-dimensional state of the system at any time t, $\mathcal{Y}(t_k) \in \mathbb{R}^r$ is the measurement at the t_k^{th} time instant, A, G_1, C, G_2, U_w and U_v are given constant matrices and B is given constant vectors of compatible dimensions. A is assumed to be a full rank matrix. $\mathcal{W}(t) \in \mathbb{R}^n$ is a standard Wiener process with increment $d\mathcal{W}(t)$ and $\mathcal{V}(k) \in \mathbb{R}^r$ is the measurement noise and zero mean, i.i.d. random vectors with identity covariance matrix $\mathcal{I}$. The standard Wiener process $\mathcal{S}_1(t) \in \mathbb{R}^n$ and the random variable $\mathcal{S}_2(t_k) \in \mathbb{R}^r$ represent the multiplicative noise. The noise signals $\mathcal{W}(t)$, $\mathcal{V}(t_k)$, $\mathcal{S}_1(t)$ and $\mathcal{S}_2(t)$ are uncorrelated with each other. The initial state is a random vector with a known mean and covariance matrix, $\mathbb{E}[\mathcal{X}(0)] = \hat{\mathcal{X}}(0)$ and $\mathbb{E}[(\mathcal{X}(0) - \hat{\mathcal{X}}(0))(\mathcal{X}(0) - \hat{\mathcal{X}}(0))^\top] = P(0)$, respectively. $\mathcal{X}(0)$, $\mathcal{W}(t)$, $\mathcal{V}(t_k)$, $\mathcal{S}_1(t)$ and $\mathcal{S}_2(t_k)$ are mutually independent.

Discretisation of process model

Let $t_{k+1} - t_k = \delta$, $k > 0$ be the uniform time interval between consecutive measurement samples. Applying the Euler scheme to (4.10) over a time interval $(t_k, t_k + \Delta)$ yields

$$\mathcal{X}(t_k^{i+1}) = \mathcal{X}(t_k^i) + (A\mathcal{X}(t_k^i) + B)\Delta + U_w\Delta W + G_1 \operatorname{diag}(\mathcal{X}(t_k^i))\Delta S_1,$$

where $t_k^i \in [t_k, t_{k+1}], i = 0, 1, \ldots, m-1$ are uniformly spaced inter-sampling times and $\Delta = \frac{\delta}{m}$. ΔW and ΔS_1 are n-dimensional Gaussian random variables with zero mean and covariance matrices $\mathbb{E}[\Delta W \Delta W^\top] = \Delta\mathcal{I}$, $\mathbb{E}[\Delta S_1 \Delta S_1^\top] = \Delta\mathcal{I}$ respectively.

In order to match exactly the first two conditional moments of $\mathcal{X}(t_k^{i+1})$ given $\mathcal{X}(t_k^i)$, we use the moment matching approach. The expression for the conditional mean of $\mathcal{X}(t_k^{i+1})$ given $\mathcal{X}(t_k^i)$ can be easily shown to be

$$\mathbb{E}[\mathcal{X}(t_k^{i+1})|\mathcal{X}(t_k^i)] = \mathcal{X}(t_k^i) + (A\mathcal{X}(t_k^i) + B))\Delta,$$

where $i = 0, 1, \ldots, m-1$ and the associated conditional covariance matrix is

$$var[\mathcal{X}(t_k^{i+1})|\mathcal{X}(t_k^i)] = U_w U_w^\top \Delta + G_1(\text{ diag}(\mathcal{X}(t_k^i)))^2 G_1^\top \Delta$$

Then

$$\mathcal{X}(t_k^{i+1}) = \tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m \tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^i))\tilde{\mathcal{S}}_1(t_k^i), \tag{4.12}$$

and

$$\mathcal{Y}(t_k^i) = C\mathcal{X}(t_k^i) + U_v \mathcal{V}(t_k^i) + G_2 \operatorname{diag}(\mathcal{X}(t_k^i))\mathcal{S}_2(t_k^i), \tag{4.13}$$

where

$$\tilde{A} = I + A\Delta,\ \tilde{B} = B\Delta, \tilde{U}_m = U_w\sqrt{\Delta},\ \tilde{G}_1 = G_1\sqrt{\Delta},$$

and $\tilde{\mathcal{W}}$, $\tilde{\mathcal{S}}_1$, $\mathcal{V}$ and $\mathcal{S}_2$ are uncorrelated, zero mean random processes with identity covariance matrices. $\mathcal{Y}(t_k^i)$ represents a pseudo-measurement which coincides with the actual measurements at $t_k, t_{k+1}, \ldots$. This puts the system in a discrete state space framework, simply with measurements missing between t_k and t_{k+1}, to which the standard discrete time filtering tools can be applied.

Approximate minimum variance filter for the continuous-discrete system with randomly delayed measurement

We consider up to N, where $N \leq m$, consecutive delayed measurements between $t_k = kT$ and $t_{k+1} = (k+1)T$. To model the delayed measurement, let us assume $p_k = \begin{bmatrix} p_{1k} & p_{2k} \cdots p_{Nk} \end{bmatrix}$ to be a vector of independent Bernoulli random variables taking values either 0 or 1 with probability

$$\begin{aligned} &P(p_{jk} = 0) = \beta, \qquad P(p_{jk} = 1) = 1 - \beta = \mathbb{E}[p_{jk}], \\ &\mathbb{E}[(p_{jk} - (1-\beta))^2] = \beta(1-\beta). \end{aligned} \tag{4.14}$$

As the delay is often small in many practical applications, the maximum delay is considered to be less than or equal to one measurement time step. Then, the measurement received at the $(t_{k+1})^{th}$ time step may actually belong to the $(t_{k+1} - j\Delta)$ time instant, $j = 0, 1, \ldots, N$. Specifically, the measurement at time t_{k+1} is as follows:

$$
\begin{aligned}
&\mathcal{Z}(t_{k+1}) = (1 - p_{1k})\mathcal{Y}(t_k^m) + p_{1k}(1 - p_{2k})\mathcal{Y}(t_k^{m-1}) + p_{1k}p_{2k}(1 - p_{3k})\mathcal{Y}(t_k^{m-2}) \\
&+ \cdots + (\prod_{j=1}^{N-1} p_{jk})(1 - p_{Nk})\mathcal{Y}(t_k^{m-(N-1)}) + (\prod_{j=0}^{N} p_{jk})\mathcal{Y}(t_k^{m-N}) \\
&= \sum_{j=0}^{N-1} \prod_{i=0}^{j} p_{ik}(1 - p_{(j+1)k})\mathcal{Y}(t_k^{m-j}) + \prod_{j=0}^{N} p_{jk}\mathcal{Y}(t_k^{m-N}), \\
&\quad t_k^i \in [t_k, t_{k+1}], \quad \mathcal{Y}(t_k^m) = \mathcal{Y}(t_{k+1}), \quad \mathcal{Y}(t_k^0) = \mathcal{Y}(t_k),
\end{aligned}
$$

with $p_{0k} := 1$. The proposed multiplicative structure is the same as in chapter 4, except that $\mathcal{Y}(t_k^{m-j})$ are not real measurements for $j = 1, 2, \cdots, m-1$.

Using $\mathbb{E}(p_{jk}) = 1 - \beta$ for $j = 1, 2, \ldots, m-1$ and using the fact that p_{jk} are independent from each other as well as from the measurement noise, we can write our one step ahead prediction of $\mathcal{Z}(t_{k+1})$ as

$$
\hat{\mathcal{Z}}(t_{k+1}|t_k) = \beta \sum_{j=0}^{N-1} (1 - \beta)^j \hat{\mathcal{Y}}(t_k^{m-j}|t_k) + (1 - \beta)^N \hat{\mathcal{Y}}(t_k^{m-N}|t_k). \tag{4.15}
$$

Combining and re-arranging,

$$
\begin{aligned}
&(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k)) \\
&= \underbrace{\sum_{j=0}^{N-1} (\prod_{i=0}^{j} p_{ik})(1 - p_{(j+1)k})(\mathcal{Y}(t_k^{m-j}) - \hat{\mathcal{Y}}(t_k^{m-j}|t_k))}_{T_1} \\
&+ \underbrace{(\prod_{j=0}^{N} p_{jk})(\mathcal{Y}(t_k^{m-N}) - \hat{\mathcal{Y}}(t_k^{m-N}|t_k))}_{T_2} + \underbrace{\sum_{j=0}^{N-1} \left((\prod_{i=0}^{j} p_{ik})(1 - p_{(j+1)k}) - \beta(1-\beta)^j \right) \hat{\mathcal{Y}}(t_k^{m-j}|t_k)}_{T_3} \\
&+ \underbrace{\left((\prod_{j=0}^{N} p_{jk}) - (1-\beta)^N \right) \hat{\mathcal{Y}}(t_k^{m-N}|t_k)}_{T_4} . \qquad (4.16)
\end{aligned}
$$

We use this re-arrangement of the *innovation* sequence $(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))$ later in the proof of our main result, given in the Appendix, for computing the covariance of the estimation error. A similar technique was also used in chapter 4.

To derive the recursive filtering equations, it is assumed that the observations are given up to time t_k and that the conditional mean of $\mathcal{X}(t_k^i)$ given $\mathcal{Z}(t_k)$, $\hat{\mathcal{X}}(t_k^i|t_k)$, is available. From this value, the conditional mean of $\mathcal{X}(t_k^{i+1})$, which provides the predictor, $\hat{\mathcal{X}}(t_k^{i+1}|t_k)$, is derived using (4.12):

$$\hat{\mathcal{X}}(t_k^{i+1}|t_k) = \tilde{A}\hat{\mathcal{X}}(t_k^i|t_k) + \tilde{B}. \qquad (4.17)$$

The predicted estimate $\hat{\mathcal{X}}(t_k^{i+1}|t_k)$ needs to be updated with the information provided by $\mathcal{Z}(t_{k+1})$, to obtain the filtered estimate. When the measurements may be randomly delayed by N sampling times, the update equation for a linear filter using N steps randomly delayed measurement is

$$\hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}) = \hat{\mathcal{X}}(t_k^{i+1}|t_k) + \bar{K}(t_{k+1}^i)(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k)), \qquad (4.18)$$

where $\hat{\mathcal{Z}}(t_{k+1}|t_k)$ is a one step ahead prediction of $\mathcal{Z}(t_{k+1})$ and the estimation error covariance matrix of the state is given by

$$\bar{P}(t_k^{i+1}|t_{k+1}) = \mathbb{E}[(\mathcal{X}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}))((\mathcal{X}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}))^\top]. \tag{4.19}$$

$\bar{P}(t_k^{i+1}|t_{k+1})$ involves expectation of terms of the form $\mathcal{X}(t_k^i)\mathcal{X}^\top(t_k^{m-j})$. To evaluate this expectation, the following result is useful.

Lemma 4. According to (4.12), the relationship between $\mathcal{X}(t_k^i)$ and $\mathcal{X}(t_k^m)$ is

$$\mathcal{X}(t_k^i) = \tilde{A}^{i-m}(\mathcal{X}(t_k^m) - \tilde{B} - \tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-1}) - \tilde{G}_1 \operatorname{diag}(\mathcal{X}_j(t_k^{m-1}))\tilde{\mathcal{S}}_1(t_k^{m-1})))+ \\ \sum_{r=1}^{m-i-1} \tilde{A}^{-(m-i-r)}(-\tilde{B} - \tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-(r+1)}) - \tilde{G}_1 \operatorname{diag}(\mathcal{X}_j(t_k^{m-(r+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(r+1)}))), \tag{4.20}$$

and the relationship between $\mathcal{X}(t_k^{m-j})$ and $\mathcal{X}(t_k^m)$ is

$$\mathcal{X}(t_k^{m-j}) = \tilde{A}^{-j}(\mathcal{X}(t_k^m) - \tilde{B} - \tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-1}) - \tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-1}))\tilde{\mathcal{S}}_1(t_k^{m-1})) \\ + \sum_{l=1}^{j-1} A^{-(j-l)}(-\tilde{B} - \tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-(l+1)}) - \tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-(l+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(l+1)})). \tag{4.21}$$

Before presenting the main result in this section, we first repeat the following lemma from chapter 3 for ease of reference (note that the notation here is different due to the two different sampling frequencies):

Lemma 5. The second order moment of $q(t_k^i)$ is given as follows:

$$q(t_k^i) = \operatorname{diag}\left(vec(P_{jj}(t_k^i|t_k)) + (\hat{\mathcal{X}}(t_k^i|t_k))^2\right), \tag{4.22}$$

where

$$q(t_k^i) = \mathbb{E}[\operatorname{diag}(\mathcal{X}(t_k^i))\operatorname{diag}(\mathcal{X}(t_k^i))^\top],$$
$$\hat{\mathcal{X}}(t_k^i|t_k) = \tilde{A}\hat{\mathcal{X}}(t_k^{i-1}|t_k) + \tilde{B},$$

and

$$\bar{P}(t_k^i|t_k) = \tilde{A}\bar{P}(t_k^{i-1}|t_k)\tilde{A}^\top + \tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{i-1})\tilde{G}_1^\top$$

Proof

$$\begin{aligned} q(t_k^i) &= \mathbb{E}[\operatorname{diag}(\mathcal{X}(t_k^i))\operatorname{diag}(\mathcal{X}(t_k^i))^\top] \\ &= \mathbb{E}[(\operatorname{diag}\mathcal{X}(t_k^i))^2]. \end{aligned} \tag{4.23}$$

On the other hand, based on the definition of the covariance matrix we have

$$\begin{aligned} &\operatorname{diag}(vec(\bar{P}_{jj}(t_k^i|t_k))) = \mathbb{E}[\operatorname{diag}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k))^2] \\ &= \mathbb{E}[(\operatorname{diag}\mathcal{X}(t_k^i))^2] - 2\mathbb{E}[\operatorname{diag}\mathcal{X}(t_k^i)]\operatorname{diag}\hat{\mathcal{X}}(t_k^i|t_k) + (\operatorname{diag}\hat{\mathcal{X}}(t_k^i|t_k))^2 \\ &= \mathbb{E}[\operatorname{diag}\mathcal{X}(t_k^i)^2] - 2\operatorname{diag}\hat{\mathcal{X}}(t_k^i|t_k)\operatorname{diag}\hat{\mathcal{X}}(t_k^i|t_k) + (\operatorname{diag}\hat{\mathcal{X}}(t_k^i|t_k))^2 \\ &= \mathbb{E}[(\operatorname{diag}\mathcal{X}(t_k^i))^2] - (\operatorname{diag}\hat{\mathcal{X}}(t_k^i|t_k))^2, \end{aligned} \tag{4.24}$$

then

$$\mathbb{E}[(\operatorname{diag}\mathcal{X}(t_k^i))^2] = \operatorname{diag}(vec(P_{jj}(t_k^i|t_k)) + (\hat{\mathcal{X}}_j(t_k^i|t_k))^2), \tag{4.25}$$

and by substituting (4.25) into (4.23), the proof of (4.22) can be completed.

The objective of this section is to find the optimum filter gain $\bar{K}(t_{k+1}^i)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(t_k^{i+1}|t_{k+1})$ of the state estimate $\hat{\mathcal{X}}(t_k^{i+1}|t_{k+1})$. Our main result in this section is given in the next theorem.

Theorem 5. For equation (4.12) and (4.13), the filter gain $\bar{K}(t_{k+1}^i)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(t_k^{i+1}|t_{k+1})$ is given by

$$\bar{K}(t_{k+1}^i) = (\sum_{j=0}^{N-1} \beta(1-\beta)^j (\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)(\tilde{A}^{-j})^\top C^\top +$$

$$\sum_{s=1}^{min\{j-1,m-i-2\}} \tilde{A}^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-s)})^\top C^\top$$

$$-\sum_{l=1}^{j-1} \tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1))})(\tilde{A}^{-(j-l)})^\top C^\top$$

$$-\tilde{A}^{i-m+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-j})^\top C^\top$$

$$-\sum_{r=1}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-r)})^\top C^\top)+$$

$$(1-\beta)^N(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)(\tilde{A}^{-N})^\top C^\top$$

$$+\sum_{s=1}^{min\{N-1,m-i-2\}} \tilde{A}^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-s)})^\top C^\top$$

$$-\sum_{l=1}^{N-1} \tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-l)})^\top C^\top$$

$$-\tilde{A}^{i-m+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-N})^\top C^\top -$$

$$\sum_{r=1}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-r)})^\top C^\top))\hat{\chi}(t_k)^{-1}$$

where

$$\bar{P}(t_k^m|t_k) = \bar{P}(t_{k+1}|t_k) = \tilde{A}\bar{P}(t_k|t_k)\tilde{A}^\top + \tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k)\tilde{G}_1^\top, \tag{4.26}$$

and $q(t_k^i)$ is as defined in (4.22) and $\hat{\chi}(t_k)$ is as defined in equation (A.54) in the Appendix.

Proof: See Appendix.
Several remarks on this result are in order.

1. If $\beta = 1$, i.e., if there is no random delay and $G_2 = 0$ i.e., if there is no multiplicative noise in the measurement, our result reduces to the filter derived in [124]. If, in addition to $\beta = 1$, we have $G_1 = 0$ and $G_2 = 0$, i.e., if there is no multiplicative noise, our filter reduces to the Kalman Bucy filter for continuous-discrete models with additive noise proposed in [3].

2. In this chapter, the differential equations that describe the process are discretised using the Euler scheme instead of Ito-Taylor expansion, since Ito-Taylor expansion leads to a nonlinear state space system for which a closed-form filter design is significantly more complex.

3. As mentioned in Chapter 3, $\bar{P}(t_{k+1}|t_k)$ is a function of $q(t_k)$ and hence of $\mathcal{Y}(t_k)$. Hence $\bar{K}(t_{k+1})$ does not converge to a time-invariant matrix and it is not possible to guarantee asymptotic stability.

4. We are seeking to minimize variance on a set of filters which have a Kalman filter-like structure: $\bar{K}(t^i_{k+1})(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))$. The filter is still not linear in $\mathcal{Z}(t_k)$, since $\bar{K}(t_{k+1})$ itself contains $\mathcal{Z}(t_k)$, as will be seen in the proof. We are not minimizing variance in the sense of finding $\mathbb{E}(\mathcal{X}(t^i_k)|\mathcal{Z}(t_k))$, which would be a very general (nonlinear) function of past and present values of $\mathcal{Z}$. Again, this fact accords with the work suggested in chapter 3. However, the filter with a Kalman filter-like structure in chapter 3 is designed to minimize the variance at the measurement sampling times (i.e., $\bar{K}(k+1)$ approximately minimizes the covariance of the estimate of state at t_{k+1}), whereas the filter in this chapter seeks to minimize the covariance of the estimate of the state at points in-between the measurement sampling instants as well as at the measurement sampling times.

4.3.3 Numerical example

We use the same model parameters as described in [77], which are: $A = \begin{bmatrix} 0 & -0.5 \\ 1 & 1 \end{bmatrix}$, $B = 0$ $U_w = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$, $C = \begin{bmatrix} -100 & 10 \end{bmatrix}$, $G_1 = \begin{bmatrix} 0.12 & 0.02 \\ 0.15 & 0.1 \end{bmatrix}$,

and $U_v = 1$. $\mathcal{W}(t)$ and $\mathcal{S}_1(t)$ are standard Wiener processes. It is worth clarifying that $\mathcal{V}(t_k)$ is random variable with zero mean and identity covariance matrix $\mathcal{I}$ and uncorrelated with $\mathcal{W}(t)$ and $\mathcal{S}_1(t)$. The initial conditions are
$\mathcal{X}(0) = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$, $\hat{\mathcal{X}}(0) = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$ and $P(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

The measurement sampling period is $\delta = 1$. We consider a sequence of $m = 10$ time steps between $t_k = k\delta$ and $t_{k+1} = (k+1)\delta$, so $\Delta = 1/10$. A difference in our simulation is that instead of using the parameters described in [77] themselves we used them to derive the discretisation parameters presented in section 4.3.2. That is,
$\tilde{A} = \begin{bmatrix} 0.9 & 0.05 \\ 0.1 & .9 \end{bmatrix}$, $\tilde{U}_w = \begin{bmatrix} -1.8974 \\ 0.3162 \end{bmatrix}$, $\tilde{G}_1 = \begin{bmatrix} 0.0379 & 0.0063 \\ 0.0474 & 0.0316 \end{bmatrix}$,

It is assumed that the observations are randomly delayed by two sample times, that is:

$$\mathcal{Z}(t_{k+1}) = (1 - p_{1k})\mathcal{Y}(t_k^m) + p_{1k}(1 - p_{2k})\mathcal{Y}(t_k^{m-1}) + p_{1k}p_{2k}\mathcal{Y}(t_k^{m-2}).$$

In order to compare the performance of the estimators, we use the RMSE criterion, as before. Consider 100 independent simulations, each with 200 data points. Denoting $\mathcal{X}^{(s)}(t_k)$, $\quad k = 1, ..., 200$ as the s^{th} set of true values of the state and $\hat{\mathcal{X}}^{(s)}(t_k|t_k)$ as the filtered state estimate at the t_k^{th} time instant for the s^{th} simulation run, the RMSE of the filter for each of the algorithms and for the i^{th} state $(i = 1, 2)$ is calculated by

$$RMSE_i(s) = \sqrt{\frac{1}{200}\sum_{k=1}^{200}(\mathcal{X}_i^{(s)}(t_k) - \hat{\mathcal{X}}_i^{(s)}(t_k|t_k))^2}, \qquad s = 1, ..., 100, i = 1, 2.$$

Then, the average of the RMSE for each of the states over 100 simulations is given by

$$AvRMSE_i = \frac{1}{100}\sum_{s=1}^{100} RMSE_i(s),\ i = 1, 2.$$

Using our estimator, we estimated the two states when the measurement is subject to at most two delays. As in chapter 3, we compare the performance of three filters: a no delay filter (i.e., $\beta = 1$), a filter designed for a single delay and a filter for two delays. The reason for this comparison is as follows. The exact delay is often not known, and it is of interest to compare the results when using a filter with an incorrect delay length. Further, it is of interest to see whether the filter with an incorrect delay still performs better than the filter with no delay at all. The results of the filter with no delay, one delay and two delays are represented by DKF_0, DKF_1 and DKF_2, respectively. Tables 4.4 and 4.5 summarize the results of this experiment. As can be seen, the new filtering algorithm with two delays outperforms the no delay filter and one delay for several different values of delays. Importantly, the results for the one delay filter are consistently better than those of the no-delay filter. Our (admittedly limited) experience indicates that de-tuning a no-delay filter for a small delay may be better than using a no-delay filter, even if the maximum number of delays is unknown.

4.4 Summary

In section 5.1 the optimal MV filter, which is linear in the current measurement, is derived for a class of continuous-discrete systems with both additive and multiplicative noise. The closed-form solution generalizes the results for

Table 4.4: Comparison of $AvRMSE_1$ for DKF_0, DKF_1 and DKF_1 for different values of β

		$\beta = 0.9$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.3$
$AvRMSE_1$	DKF_0	13.5072	15.7805	17.0713	18.5921
	DKF_1	9.6485	10.1942	10.4937	11.1298
	DKF_2	8.8444	9.0650	9.1764	9.5543

Table 4.5: Comparison of $AvRMSE_2$ for DKF_0, DKF_1 and DKF_2 for different values of β

		$\beta = 0.9$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.3$
$AvRMSE_2$	DKF_0	3.0353	3.5572	3.8510	4.1912
	DKF_1	2.6837	2.6736	2.7168	2.8350
	DKF_2	2.2916	2.4312	2.5781	2.7523

MV filtering for additive-multiplicative noise cases in [77]. The continuous time dynamics are discretized using the Euler scheme with a smaller time step than the measurement sampling time. The results of this section were applied to simulated linear system with additive-multiplicative noises. Our numerical experiments indicate that the CDF outperforms DDF.

Moreover, we present an approximate MV filter for a class of continuous discrete time systems with both additive and multiplicative noise when the measurement might be delayed randomly in section 5.2. A numerical example has been provided to illustrate the effectiveness of the proposed design approach. Our numerical experiments indicate that the proposed algorithm outperforms an implementation which ignores delays for a range of delay probabilities. Further, it appears that, in the presence of an unknown number of delays, using a single delay filter might still be better than using a filter with no delays.

Chapter 5

Approximate minimum variance filter under random delays III: nonlinear systems with additive noise

5.1 Introduction

In the previous chapters we considered a *linear* system with an interpretation of multiplicative uncertainty as stochastic uncertainty in parameters. It is ,however, an accepted fact that, in many applications of interest, the system dynamics and the observation equation are nonlinear This chapter is concerned with the problem of state estimation for nonlinear discrete time systems. Unlike a linear Gaussian discrete-time system, the optimal recursive solution to the state estimation problem in nonlinear systems is usually not available in closed form. In the past few years, considerable attention has been devoted to the nonlinear filtering problem. A series of suboptimal approaches have been developed in the literature to solve the nonlinear filtering problem, including the the EKF [10], the UKF [11], EnKF [12] and the QKF [13]. These are reviewed earlier in chapter 2.

An extensive theory and a number of algorithms have been developed for filtering in nonlinear systems using linearisation, *e.g.* [126]. If computational time is not an issue, there are powerful sampling-based particle filtering algorithms [34] available which are guaranteed to converge and outperform the suboptimal methods if a sufficient number of samples are generated at each time step. However, linearisation-based filters are still often preferred in real-time systems where sampling tends to be an expensive operation.

A separate strand of literature considers randomly delayed measurements, which are frequently encountered in many practical applications, such as target tracking, communication, signal processing, control and transportation [155, 156, 157] and [158]. In order to make sure that the filtering error dynamics converge, possible measurement delays should be taken into account in designing filters. In the past few years, many results have been reported in the literature on the estimation of the state of nonlinear discrete-time systems with randomly delayed observations; see, e.g., [119]. However, the literature on nonlinear filtering from delayed observations is less extensive. In [159], different approximations of the statistics of a nonlinear transformation of a random vector are used to investigate the filtering problem for a class of nonlinear stochastic systems with randomly delayed observations, where the possible delay is restricted to a single step and characterized by a set of Bernoulli variables. This work is extended in [122] to address the case when the measurements might be delayed randomly by one or two sampling steps. For the same model, a Gaussian approximation filter is derived in [121]. In [160], an EKF algorithm is designed which provides optimal estimates of interconnected network states when some or all the measurements are delayed. In [161], the filtering problem for a class of nonlinear discrete time stochastic systems with delayed states is investigated.

To the best of my knowledge, the problem of state estimation in nonlinear systems with random delays has not been fully investigated. This chapter is concerned with the problem of state estimation for a nonlinear discrete time system with a single random delay in measurement. Unlike chapters 3-4, the state space model is nonlinear and requires linearisation at each time step with the covariance of multiplicative noise, which is used as a proxy for the magnitude of the linearisation error, varies at each time step. A heuristic for choosing the parameters representing the magnitude of linearisation error is suggested, and tested through a numerical example. The material presented in this chapter has been published in [162]

5.2 System model and problem formulation

Our aim is to address the problem of the state estimation for nonlinear discrete time systems with additive noise where the measurement might be delayed randomly by one sample time when the Bernoulli random variables describing the delayed observations, with values one or zero indicating whether or not the measurement is delayed.

Consider a class of discrete-time nonlinear stochastic systems with additive noise where the measurement might be delayed by one time step. Using a notation similar to that in the previous chapters, the model is described by the following state and measurement equations:

$$\mathcal{X}(k+1) = f(\mathcal{X}(k)) + U_w\mathcal{W}(k), \tag{5.1}$$

$$\mathcal{Y}(k) = h(\mathcal{X}(k)) + U_v\mathcal{V}(k), \tag{5.2}$$

$$\mathcal{Z}(k) = (1-p_k)\mathcal{Y}(k) + p_k\mathcal{Y}(k-1), \tag{5.3}$$

where $\mathcal{X}(k) \in \mathbb{R}^n$ is the state vector at a time k to be estimated, $\mathcal{Y}(k) \in \mathbb{R}^r$ is the measurement vector at time k and $\mathcal{Z}(k) \in \mathbb{R}^r$ is the one step randomly delayed measurement equation alternatively referred to as "received measurement", p_k denotes the Bernoulli random variable at each time k (binary switching sequence taking the values 0 or 1) with a known distribution $\mathbb{P}(p_k = 0) = \beta$ and $\mathbb{P}(p_k = 1) = 1-\beta$ and p_k are uncorrelated with other random variables. U_w and U_v are given deterministic matrices. $\mathcal{W}(k) \in \mathbb{R}^n$ and $\mathcal{V}(k) \in \mathbb{R}^r$ are the process noise and the measurement noise, respectively. The nonlinear functions $f(\mathcal{X}(k))$ and $h(\mathcal{X}(k))$ are at least twice differentiable, with a known form. The work of this chapter is carried out based on the following assumptions:

Assumption 1: The noise signals $\mathcal{W}(k)$ and $\mathcal{V}(k)$ are zero mean, i.i.d. random vectors with identity covariance matrix $\mathcal{I}$ and mutually uncorrelated.

Assumption 2: The initial state is a random vector with a known mean and covariance matrix, $\mathbb{E}[\mathcal{X}(0)] = \hat{\mathcal{X}}(0)$ and $\mathbb{E}[(\mathcal{X}(0) - \hat{\mathcal{X}}(0))(\mathcal{X}(0) - \hat{\mathcal{X}}(0))^\top] = P(0)$, respectively. $\mathcal{X}(0)$, $\mathcal{W}(k)$ and $\mathcal{V}(k)$ are mutually independent.

The approximated conditional mean of $\mathcal{X}(k+1)$, which provides the predictor $\hat{\mathcal{X}}(k+1|k)$, is derived using (5.1):

$$\hat{\mathcal{X}}(k+1|k) = f(\hat{\mathcal{X}}(k|k)). \tag{5.4}$$

For brevity of notation, an expression $LL^\top$ will sometimes be denoted as $(L)(\star)^\top$, where L is a matrix-valued or vector-valued expression and where there is no risk of confusion.

The update equation for a nonlinear filter using one step randomly delayed measurements is

$$\hat{\mathcal{X}}(k+1|k+1) = \hat{\mathcal{X}}(k+1|k) + \bar{K}(k+1)(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k)), \tag{5.5}$$

and the estimation error covariance matrix is given by

$$
\begin{aligned}
&\bar{P}(k+1|k+1) = \mathbb{E}[(\mathcal{X}(k+1) - \hat{\mathcal{X}}(k+1|k+1))(\star)^\top] \\
&= \mathbb{E}[(f(\mathcal{X}(k)) + U_w \mathcal{W}(k) - (f(\hat{\mathcal{X}}(k|k)) + \bar{K}(k+1)(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k))))(\star)^\top]
\end{aligned}
\tag{5.6}
$$

where $\quad \mathcal{Z}(k+1) = (1-p_k)(h(\mathcal{X}(k+1)) + U_v \mathcal{V}(k+1)) + p_k(h(\mathcal{X}(k)) + U_v \mathcal{V}(k)).$ (5.7)

Note that the update equation is linear in the current (possibly delayed) measurement $\mathcal{Z}(k+1)$, which is in keeping with our approach in chapters 4 and 5. By using the Taylor series expansion around $\hat{\mathcal{X}}(k|k)$, we linearize $f(\mathcal{X}(k))$ and $h(\mathcal{X}(k))$ as follows:

$$
f(\mathcal{X}(k)) = f(\hat{\mathcal{X}}(k|k)) + A(k)\tilde{\mathcal{X}}(k|k) + o(|\tilde{\mathcal{X}}(k|k)|),
$$

$$
h(\mathcal{X}(k+1)) = h(\hat{\mathcal{X}}(k+1|k)) + C(k+1)\tilde{\mathcal{X}}(k+1|k) + g(|\tilde{\mathcal{X}}(k+1|k)),
$$

where

$$
\begin{aligned}
A(k) &= \frac{\partial f(\mathcal{X}(k))}{\partial \mathcal{X}(k)}|_{\mathcal{X}(k)=\hat{\mathcal{X}}(k|k)}, \\
C(k+1) &= \frac{\partial h(\mathcal{X}(k+1))}{\partial \mathcal{X}(k+1)}|_{\mathcal{X}(k+1)=\hat{\mathcal{X}}(k+1|k)} \\
\tilde{\mathcal{X}}(k+i|k) &= \mathcal{X}(k+i) - \hat{\mathcal{X}}(k+i|k) \quad i = 0, 1.
\end{aligned}
\tag{5.8}
$$

In [112], $o(|\tilde{\mathcal{X}}(k|k)|)$ are characterized as

$$
o(|\tilde{\mathcal{X}}(k|k)|) = B(k)N(k)L(k)\tilde{\mathcal{X}}(k|k),
$$

where $B(k)$ are bounded problem-dependent scaling matrices, $L(k)$ provides an extra degree of freedom to tune the filter and $N(k)$ are unknown time-varying matrices accounting for the linearisation errors of the dynamical model and satisfies

$$
N(k)N(k)^\top \leq I.
$$

In the work presented here, we characterise the linearisation error $o(|\tilde{\mathcal{X}}(k|k)|)$ and $g(|\tilde{\mathcal{X}}(k+1|k)|)$ as stochastic perturbations which are linear in $\tilde{\mathcal{X}}(k|k)$ and $\tilde{\mathcal{X}}(k+1|k)$:

$$\begin{aligned} o(|\tilde{\mathcal{X}}(k|k)|) &= Q_1(k)\tilde{\mathcal{X}}(k|k)\mathcal{R}_1(k), \\ g(|\tilde{\mathcal{X}}(k+1|k)|) &= Q_2(k+1)\tilde{\mathcal{X}}(k+1|k)\mathcal{R}_2(k+1). \end{aligned}$$

This gives us an approximate equivalent linear system with additive-multiplicative noise:

$$\mathcal{X}(k+1) = f(\hat{\mathcal{X}}(k|k)) + A(k)\tilde{\mathcal{X}}(k|k) + Q_1(k)\tilde{\mathcal{X}}(k|k)\mathcal{R}_1(k) + U_w\mathcal{W}(k), \tag{5.9}$$

$$\begin{aligned} \mathcal{Y}(k+1) = {}& h(\hat{\mathcal{X}}(k+1|k)) + C(k+1)\tilde{\mathcal{X}}(k+1|k) \\ & + Q_2(k+1)\tilde{\mathcal{X}}(k+1|k)\mathcal{R}_2(k+1) + U_v\mathcal{V}(k+1), \end{aligned} \tag{5.10}$$

where $\mathcal{R}_1(k) \in \mathbb{R}^n$ and $\mathcal{R}_2(k+1) \in \mathbb{R}^r$ are zero mean, i.i.d. random vectors with identity covariance matrix $\mathcal{I}$ and are mutually independent with the initial state and other noise signals. $Q_1(k) \in R^n$ and $Q_2(k+1) \in R^r$ describe the effect of higher-order terms in the Taylor series in terms of parameter uncertainties. In particular, the matrices $Q_1(k)$ and $Q_2(k)$ appear in the computation of covariance matrix of the state estimate only in the form $Q_1(k)\Lambda(k)Q_1(k)$ and $Q_2(k)\Lambda Q_2(k)$, where $\Lambda(k)$ is a positive definite matrix; see equations (6.11) and (A.66). The justification of characterising deterministic Taylor series truncation error by stochastic multiplicative noise is as follows. Firstly, we are typically interested in filter tracking performance over a period of time, e.g., as measured by the root mean squared error, and treating the error as stochastic can be advantageous if it yields a closed-form result (as is the case here). Secondly, as demonstrated in the numerical example presented, the *size* of the stochastic uncertainty representing the linearisation error can be used as a tuning parameter for the linearized filter in order to improve the filtering performance.

The objective of this section is to find the optimum filter gain $\bar{K}(k+1)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(k+1|k+1)$ of the state estimate $\hat{\mathcal{X}}(k+1|k+1)$ for the approximate linear system given by (5.7), (5.8), (5.9) and (5.10). Our main result in this section is given in the next theorem.

Theorem 6. For equations (5.7)-(5.10), the filter gain $\bar{K}(k+1)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(k+1|k+1)$ is given by

$$\begin{aligned}
&\bar{K}(k+1) = (\beta\bar{P}(k+1|k)C(k+1)^\top + (1-\beta)A(k)\bar{P}(k|k)C(k)^\top) \\
&\times [\beta(C(k+1)\bar{P}(k+1|k)C(k+1)^\top + Q_2(k+1)\bar{P}(k+1|k)Q_2(k+1)^\top + U_vU_v^\top)+ \\
&(1-\beta)(C(k)\bar{P}(k|k)C(k)^\top + Q_2(k)\bar{P}(k|k)Q_2(k)^\top + U_vU_v^\top) + \beta(1-\beta)(\widetilde{\psi}_0(k+1) \\
&+ \widetilde{\psi}_1(k+1)) - \beta(1-\beta)(\psi_0(k+1)\psi_1(k+1)^\top + \psi_1(k+1)\psi_0(k+1)^\top)]^{-1}
\end{aligned} \tag{5.11}$$

where $\bar{P}(k+1|k)$ and $\psi_i(k+1)$, $\widetilde{\psi}_i(k+1)$, $\quad i=0,1$ are as defined in (A.62) and (A.63), respectively, in the Appendix.

Proof: See Appendix.
Remarks:

- It is easy to verify that setting $\beta = 1$ and $Q_1(k) = Q_2(k) = 0$ gives the familiar Extended Kalman filter for the delay-free case.
- As in chapters 3 and 4, note that $\bar{P}(k+1|k)$ is a function of $\hat{\mathcal{X}}(k|k)$ and hence of $\mathcal{Y}(k)$. $\bar{K}(k+1)$ thus cannot converge to a time-invariant matrix as $k \to \infty$, and asymptotic stability cannot be guaranteed in this case.

5.3 Numerical Example

To test the accuracy of the new algorithm, the following univariate non-stationary growth model is considered,

$$\mathcal{X}(k+1) = a\mathcal{X}(k) + b\frac{\mathcal{X}(k)}{1+\mathcal{X}(k)^2} + d\cos(1.2k) + U_w\mathcal{W}(k),$$
$$\mathcal{Y}(k) = \frac{\mathcal{X}(k)^2}{20} + U_v\mathcal{V}(k),$$

where $\mathcal{V}(k)$ and $\mathcal{W}(k)$ are i.i.d. random variables with zero mean and unit variance. This model has been previously used in [9]. We use the parameters $a = 0.5$, $b = 1$, $d = 8$, $U_w = 0.1$ and $U_v = 0.1$. Initial conditions are $\mathcal{X}(0) = 1$, $\hat{\mathcal{X}}(0) = 0$ and $P(0) = 0.1$. In equations (5.9) and (5.10), we use $Q_1(k) = \gamma\,\mathrm{trace}\,(A(k))$ and $Q_2(k) = \gamma\,\mathrm{trace}\,(C(k))$, where γ is our tuning parameter that expresses the linearisation error as a percentage of linearized parameters. As the model has strong nonlinearities, we expect that using a large non-zero gamma might improve the performance.

In order to evaluate the efficiency of the estimators, we use the root mean square error (RMSE) criterion. Consider 100 independent simulations, each with 200 data points. Denoting $\mathcal{X}^{(s)}(k), \quad k = 1, ..., 200$ as the s^{th} set of true values of the state and $\hat{\mathcal{X}}^{(s)}(k|k)$ as the filtered state estimate at time k for the s^{th} simulation run, the RMSE is calculated by

$$RMSE(s) = \sqrt{\frac{1}{200}\sum_{k=1}^{200}(\mathcal{X}^{(s)}(k) - \hat{\mathcal{X}}^{(s)}(k|k))^2},$$
$$s = 1, ..., 100.$$

Then, the average of RMSE of the state over 100 simulations is given by

$$AvRMSE = \frac{1}{100}\sum_{s=1}^{100} RMSE(s).$$

We perform two experiments for this system with different levels of linearisation error γ. Firstly, to isolate the improvement in performance even in the absence of delay, we conduct a delay-free experiment (i.e., with $\beta = 1$) and compare the performance of a filter with pure linearisation or EKF (i.e., $\gamma = 0$) with a filter with different values of γ (i.e., $\gamma = 0.25, 0.5, 0.75$). Table 5.1 summarizes the results of this experiment. As can be seen, the filtering algorithm with the linearisation error accounted for in terms of multiplicative uncertainty outperforms the filtering algorithm with pure linearisation (i.e., $\gamma = 0$). Further, improvement in the performance of filter becomes more pronounced as γ increases.

Table 5.1: Comparison of $AvRMSE$ for different values of γ

	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.5$	$\gamma = 0.75$
$AvRMSE$	0.1169	0.1168	0.1166	0.1164

In the second numerical experiment, we compared a one delay filter with pure linearisation (i.e., $\gamma = 0$) with a one delay filter using interpretation of multiplicative noise in terms of linearisation error. We used different levels of multiplicative uncertainty as a proxy for linearisation error (i.e., $\gamma = 0.25, 0.5, 0.75$) and different values of β (i.e., $\beta = 0.9, 0.5, 0.1$). Recall that the actual probability of receiving a delayed measurement is $1 - \beta$. Table 5.2 summarizes the results of this experiment. As can be seen, the filtering algorithm with uncertainties outperforms the filtering algorithm with pure linearisation (i.e., $\gamma = 0$) in all cases. The increase in the error with increase in delay probability is in keeping with the comments in [112]. Note that increasing γ to a value larger than 1 makes the filter unstable.

5.4 Summary

In this chapter, an approximate MV filter is discussed for a class of nonlinear discrete time systems with additive noise and a random delay. The chap-

Table 5.2: Comparison of $AvRMSE$ for different values of β and γ

		$\beta = 0.9$	$\beta = 0.5$	$\beta = 0.1$
$AvRMSE$	$\gamma = 0$	0.1170	0.1216	0.1402
	$\gamma = 0.25$	0.1169	0.1215	0.1401
	$\gamma = 0.5$	0.1168	0.1214	0.1399
	$\gamma = 0.75$	0.1167	0.1213	0.1398

ter makes two distinct contributions. Firstly, it generalizes the closed-form solution for MV filtering for linear systems in [56]. We have used a novel approach of modelling the linearisation error as multiplicative noise that in essence de-tunes the EKF to account for this noise or the linearisation error. Secondly, we extend the results to cope with a random delay of a single time step. Our numerical experiment indicates that the proposed filtering algorithm can be used to improve filtering performance as measured by root mean squared error when linearized dynamics is used for filter design.

Chapter 6

Conclusions and directions for future research

This chapter concludes the thesis by summarizing the main contributions and suggests directions for future research.

6.1 Summary of contributions

In this thesis, the problem of MV filter has been investigated for different class of systems i.e. linear discrete systems, continuous-discrete time linear systems and nonlinear discrete systems. Specifically, we propose a various of an approximate filtering approaches considering of the both additive and multiplicative noise. Then, we extended these results to deal with situations when the measurement is delayed by one or more sample times. The proposed filtering algorithms are obtained by using innovation analysis approach.

In chapter 3, an approximate MV filter for a class of discrete time systems with both additive and multiplicative noise is investigated. The utility of the proposed algorithm is tested on two examples, one a two-factor extension of the Vasicek interest rate model and one based on parameters estimated from real-world data. It is also shown that the proposed filtering algorithm clearly outperforms the Kalman filter, which ignores parameter uncertainties. For the same system, a new approximate MV filter with randomly delayed observations is proposed. The number of sample times by which the observation is delayed is considered to be uncertain. We model the observations delayed by up to N sample times by using N Bernoulli random variables with values 0 or 1, and derive a closed-form expression for the proposed MV filter for this system which are linear in the current measurement while being nonlinear in one or more past measurements. The utility of the proposed filtering algorithm is demonstrated through comprehensive numerical experiments. Our numerical experiments indicate that the proposed algorithm outperforms an implementation which ignores delays, for a range of delay probabilities.

A filter which minimizes the variance of state estimates at points in-between the measurement sampling instants can improve the result compared with the filter proposed in chapter 3. This idea has been investigated in chapter 4. An approximate MV filter for a class of continuous discrete systems with both additive and multiplicative noise is investigated. The Euler scheme followed by conditional moment matching is used to transform SDEs in the process equation into a discrete model on a timescale which is finer than the measurement timescale. We test the performance of our new filter *i.e.* CDF on simulated numerical examples and compare the results with the DDF which ignores the state behaviour in-between the measurement samples. The results show that the CDF outperforms the DDF in all cases examined. The results proposed in Section 4.1 have also been extended to deal with situations when the measurement is delayed by one or more sample times. The number of sample times by which the observation is delayed is considered to be uncertain and a fraction of the measurement sample time. As in Section 3.1, the Euler scheme is used to transform the process into a discrete time state space system with a higher sampling frequency than the measurement frequency. Closed-form solution for variance minimizing filter which is linear in the current measurement while being nonlinear in one or more past measurements is obtained for this system. The utility of the proposed filtering algorithm is demonstrated through numerical experiments. Results illustrate the improved accuracy achieved by the new filters when compared to the filter which ignores delays.

The result presented in chapter 3 is extended to deal with nonlinear state space models of the discrete time systems with additive noise where the measurement might be delayed randomly by one sample time in chapter 5. As in the previous chapters, we model the observations delayed by one sample time by using Bernoulli random variables with values 0 or 1. We model the linearisation error for system in terms of multiplicative noise and then derive a closed-form expression for the MV filter for a system with multiplicative noise and random delay. The performance of the proposed filter with different levels of linearisation error and EKF has been tested on numerical example. The results indicate that the proposed filtering algorithm can be used to improve the filtering performance as measured by root mean squared error when a linearized filter is used.

6.2 Suggestions for future research

The results presented in this thesis and outlined in the previous section lead in several directions that could be studied in the future.

1. The MV filter for systems with randomly delayed observations and additive-multiplicative noise is derived in this thesis. It would be interesting to extend these results to deal with packet dropouts and missing measurements.

2. The MV filter for systems with randomly delayed observations and additive-multiplicative noise is derived in this thesis with uncorrelated noises. It might also be of interest to extend these results to correlated noises.

3. In chapter 4 the transition equation is represented in continuous time, while the measurement equation is made at discrete instances of time. It might be of interest to consider the measurement equation in the continuous time as well.

4. The methods presented in chapters 3 and 4 can be used to generate a proposal density for a PF designed to deal with systems having random delays and additive-multiplicative noise.

6.3 Summary

To conclude, An approximate minimum variance filter for different class of systems (i.e. linear discrete systems, continuous-discrete time linear systems and nonlinear discrete systems) with both additive and multiplicative noise have been developed and have been tested on several numerical examples. Then, a new state estimation algorithm for latent state variables is proposed nd have been tested on several numerical examples for these class of systems with both additive and multiplicative noise, where the measurement might be delayed randomly by one or more sample times. Also, closed-form solution for variance minimizing filter which is linear in the current measurement while being nonlinear in one or more past measurements is obtained in all the cases. However, the proposed filters can be applied to state estimation of a system in which the estimator consists of sensors connected through communication networks and many real-world applications where measurements are delayed randomly and there is parametric uncertainty to be accounted for.

Appendix A

Appendix

Proof of Theorem 1

The filtering estimates of the state covariance is obtained by combining the equations (3.1)-(3.6) as follows. For brevity of notation, an expression $LL^\top$ will sometimes be denoted as $(L)(\star)^\top$, where L is a matrix-valued expression and where there is no risk of confusion. The proof below is a straightforward modification of a similar proof in [77] and reproduced here for the sake of completeness. The estimation error covariance matrix at time $k+1$ can be written as

$$
\begin{aligned}
&\bar{P}(k+1|k+1) = \mathbb{E}[(\mathcal{X}(k+1) - \hat{\mathcal{X}}(k+1|k+1))(\star)^\top] \\
&= \mathbb{E}[((A\mathcal{X}(k) + B + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k)) \\
&\ - (A\hat{\mathcal{X}}(k|k) + B + \bar{K}(k+1)(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^\top] \\
&= \mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(\star)^\top A^\top] + U_w U_w^\top + G_1 q(k|k) G_1^\top + \bar{K}(k+1)(\mathbb{E}[(\mathcal{Y}(k+1) \\
&\ - \hat{\mathcal{Y}}(k+1|k))(\star)^\top]\bar{K}(k+1)^\top - \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + \\
&G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))^\top]\bar{K}(k+1)^\top - \bar{K}(k+1)\mathbb{E}[(\mathcal{Y}(k+1) \\
&\ - \hat{\mathcal{Y}}(k+1|k))(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))^\top]. \qquad \text{(A.1)}
\end{aligned}
$$

Next, we need the following covariance term in evaluating $\bar{P}(k+1|k+1)$:

$$\mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^\top] = CA\bar{P}(k|k)A^\top C^\top + CU_wU_w^\top C^\top + CG_1q(k|k)G_1^\top C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top. \quad \text{(A.2)}$$

We also need to evaluate some cross covariance terms, whose expressions are derived next:

$$\begin{aligned}
&\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + Uw\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))^\top] \\
&= \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(CA\mathcal{X}(k) + CU_w\mathcal{W}(k) + \\
&CG_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k) + U_v\mathcal{V}(k+1) + G_2 \operatorname{diag}(\mathcal{X}(k+1))\mathcal{S}_2(k+1) - CA\hat{\mathcal{X}}(k|k)))^\top] \\
&= \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))^\top A^\top C^\top)] + U_wU_w^\top C^\top + G_1q(k|k)G_1^\top C^\top \\
&= A\bar{P}(k|k)A^\top C^\top + U_wU_w^\top C^\top + G_1q(k|k)G_1^\top C^\top. \qquad \text{(A.3)}
\end{aligned}$$

Further,

$$\begin{aligned}
&\mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + Uw\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))^\top] \\
&= CA\bar{P}(k|k)A^\top + CU_wU_w^\top + CG_1q(k|k)G_1^\top. \qquad \text{(A.4)}
\end{aligned}$$

Substituting (A.2), (A.3) and (A.4) in (A.1), we have

$$\begin{aligned}
&\bar{P}(k+1|k+1) = A\bar{P}(k|k)A^\top + U_wU_w^\top + G_1q(k|k)G_1^\top + \bar{K}(k+1)(CA\bar{P}(k|k)A^\top C^\top \\
&+ CU_wU_w^\top C^\top + CG_1q(k|k)G_1^\top C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)\bar{K}(k+1)^\top \\
&- (A\bar{P}(k|k)A^\top C^\top + U_wU_w^\top C^\top + G_1q(k|k)G_1^\top C^\top)\bar{K}(k+1)^\top - \\
&\bar{K}(k+1)(CA\bar{P}(k|k)A^\top + CU_wU_w^\top + CG_1q(k|k)G_1^\top) \\
&= \bar{P}(k+1|k) + \bar{K}(k+1)(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)\bar{K}(k+1)^\top \\
&- \bar{P}(k+1|k)C^\top\bar{K}(k+1)^\top - \bar{K}(k+1)C\bar{P}(k+1|k), \qquad \text{(A.5)}
\end{aligned}$$

where $\bar{P}(k+1|k)$ is as defined in (3.13). To find the value of $\bar{K}(k+1)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(k+1|k+1)$ we differentiate the trace of the above expression with respect to matrix $\bar{K}(k+1)$ and set the derivative to zero.

$$\frac{\partial tr\bar{P}(k+|k+1)}{\partial\bar{K}(k+1)} = -2\bar{P}(k+1|k)C^\top + 2\bar{K}(k+1)[C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top]. \quad \text{(A.6)}$$

Setting $\frac{\partial tr\bar{P}(k+|k+1)}{\partial \bar{K}(k+1)} = 0$ leads to

$$\bar{K}(k+1) = \bar{P}(k+1|k)C^{\top}[C\bar{P}(k+1|k)C^{\top} + U_v U_v^{\top} + G_2 q(k+1|k)G_2^{\top}]^{-1}, \tag{A.7}$$

which is the required expression. ∎

Proof of Theorem 2

The proof is on the same lines as the proof of Theorem 1. The filtering estimates of the covariance matrix is obtained by combining the equations (3.23)-(3.25) as follows. The estimation error covariance matrix at time $k+1$ can be written as

$$\begin{aligned}
&\bar{P}(k+1|k+1) = \mathbb{E}[(\mathcal{X}(k+1) - \hat{\mathcal{X}}(k+1|k+1))(\star)^{\top}] \\
&= \mathbb{E}[(A\mathcal{X}(k) + B + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k) - (A\hat{\mathcal{X}}(k|k) + B \\
&+ \bar{K}(k+1)((1-p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)) + p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)) \\
&+ ((1-p_{k+1}) - \beta)\hat{\mathcal{Y}}(k+1|k) + (p_{k+1} - (1-\beta))\hat{\mathcal{Y}}(k|k))))(\star)^{\top}] \\
&= \mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(\star)^{\top}] + U_w U_w^{\top} + G_1 q(k|k) G_1^{\top} + \bar{K}(k+1)(\mathbb{E}[(1-p_{k+1})^2]\times \\
&\mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^{\top}] + \mathbb{E}[(p_{k+1})^2]\mathbb{E}[(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k))(\star)^{\top}] + \\
&\mathbb{E}[((1-p_{k+1}) - \beta)^2](\hat{\mathcal{Y}}(k+1|k))(\star)^{\top} + \mathbb{E}[(p_{k+1} - (1-\beta))^2](\hat{\mathcal{Y}}(k|k))(\star)^{\top} \\
&+ \mathbb{E}[(1-p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)))^{\top}] \\
&+ \mathbb{E}[(p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)))((1-p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)))^{\top}] \\
&+ \mathbb{E}[((1-p_{k+1}) - \beta)(p_{k+1} - (1-\beta))(\hat{\mathcal{Y}}(k+1|k)\hat{\mathcal{Y}}(k|k)^{\top} + \hat{\mathcal{Y}}(k|k)\hat{\mathcal{Y}}(k+1|k)^{\top})])\times \\
&\bar{K}(k+1)^{\top} - \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k)) \\
&((1-p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)) + p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)) + ((1-p_{k+1}) - \beta)\hat{\mathcal{Y}}(k+1|k) \\
&+ (p_{k+1} - (1-\beta))\hat{\mathcal{Y}}(k|k))^{\top}]\bar{K}(k+1)^{\top} - \bar{K}(k+1)\mathbb{E}[((1-p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)) + \\
&p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)) + ((1-p_{k+1}) - \beta)\hat{\mathcal{Y}}(k+1|k) + (p_{k+1} - (1-\beta))\hat{\mathcal{Y}}(k|k)) \\
&(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))^{\top}].
\end{aligned} \tag{A.8}$$

Next, we need the following covariance terms in evaluating $\bar{P}(k+1|k+1)$:

$$
\begin{aligned}
&\mathbb{E}[(1-p_{k+1})^2] = \mathbb{P}[1-p_{k+1}=0](0)^2 + \mathbb{P}[1-p_{k+1}=1](1)^2 \\
&= \mathbb{P}[p_{k+1}=1](0)^2 + \mathbb{P}[p_{k+1}=0](1)^2 = \beta, \\
&\mathbb{E}[(p_{k+1})^2] = \mathbb{P}[p_{k+1}=0](0)^2 + \mathbb{P}[p_{k+1}=1](1)^2 = 1-\beta, \\
&\mathbb{E}[(1-p_{k+1})-\beta)^2] = \mathbb{E}[(1-p_{k+1})^2 - 2(1-p_{k+1})\beta + \beta^2] = \beta - 2\beta^2 + \beta^2 \\
&= \beta - \beta^2 = \beta(1-\beta), \\
&\mathbb{E}[(1-p_{k+1})p_{k+1}] = \mathbb{E}[p_{k+1} - p_{k+1}^2] = \beta - \beta = 0, \\
&\mathbb{E}[((1-p_{k+1})-\beta)(p_{k+1}-(1-\beta))] = \mathbb{E}[(1-p_{k+1})p_{k+1} - (1-p_{k+1})(1-\beta) \\
&- p_{k+1}\beta + \beta(1-\beta)] = 0 - \beta(1-\beta) - \beta(1-\beta) + \beta(1-\beta) = -\beta(1-\beta),
\end{aligned}
\tag{A.9}
$$

We can also derive the following easily:

$$
\begin{aligned}
&\mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^\top] = \mathbb{E}[(C\mathcal{X}(k+1) + D + U_v\mathcal{V}(k+1) \\
&+ G_2\,\mathrm{diag}(\mathcal{X}(k+1))\mathcal{S}_2(k+1) - (C\hat{\mathcal{X}}(k+1|k) + D))(\star)^\top] \\
&= C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2 q(k+1|k)G_2^\top, \\
&\mathbb{E}[(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k))(\star)^\top] \\
&= \mathbb{E}[(C\mathcal{X}(k) + D + U_v\mathcal{V}(k) + G_2\,\mathrm{diag}(\mathcal{X}(k))\mathcal{S}_2(k) - (C\hat{\mathcal{X}}(k|k) + D))(\star)^\top] \\
&= C\bar{P}(k|k)C^\top + U_vU_v^\top + G_2 q(k|k)G_2^\top,
\end{aligned}
\tag{A.10}
$$

We also need to evaluate some cross covariance terms, whose expressions are derived next:

$$
\begin{aligned}
&\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1\,\mathrm{diag}(\mathcal{X}(k))\mathcal{S}_1(k))((1-p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)) \\
&+ p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)) + ((1-p_{k+1}) - \beta)\hat{\mathcal{Y}}(k+1|k) + (p_{k+1} - (1-\beta))\hat{\mathcal{Y}}(k|k))^\top] \\
&= \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1\,\mathrm{diag}(\mathcal{X}(k))\mathcal{S}_1(k)) \\
&((1-p_{k+1})(CA\mathcal{X}(k) + CU_w\mathcal{W}(k) + CG_1\,\mathrm{diag}(\mathcal{X}(k))\mathcal{S}_1(k) + U_v\mathcal{V}(k+1) + \\
&G_2\,\mathrm{diag}(\mathcal{X}(k+1))\mathcal{S}_2(k+1) - CA\hat{\mathcal{X}}(k|k)))^\top] + \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) +
\end{aligned}
$$

$$G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(p_{k+1}(C\mathcal{X}(k) + U_v\mathcal{V}(k) + G_2 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_2(k) - C\hat{\mathcal{X}}(k|k)))^\top]$$
$$+ \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))$$
$$(((1 - p_{k+1}) - \beta)\psi_0(k+1) + (p_{k+1} - (1-\beta))\psi_1(k+1))^\top]$$
$$= \mathbb{E}[1 - p_{k+1}]\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))^\top A^\top C^\top)] + U_wU_w^\top C^\top$$
$$+ G_1q(k)G_1^\top C^\top + \mathbb{E}(p_{k+1})(\mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(C(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))^\top]$$
$$= \beta(A\bar{P}(k|k)A^\top C^\top + U_wU_w^\top C^\top + G_1q(k|k)G_1^\top C^\top) + (1-\beta)A\bar{P}(k|k)C^\top$$
$$= \beta\bar{P}(k+1|k)C^\top + (1-\beta)AP(k|k)C^\top \tag{A.11}$$

Similarly,

$$\mathbb{E}[((1 - p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)) + p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)) + ((1 - p_{k+1}) - \beta)\psi_0(k+1)$$
$$+ (p_{k+1} - (1-\beta))\psi_1(k+1))(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}_j(k))\mathcal{S}_1(k))^\top]$$
$$= \beta C\bar{P}(k+1|k) + (1-\beta)C\bar{P}(k|k)A^\top, \tag{A.12}$$

Substituting (A.9)- (A.12) in (A.8), we have

$$\bar{P}(k+1|k+1) = \bar{P}(k+1|k) + \bar{K}(k+1)(\beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)$$
$$+ (1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top + G_2q(k|k)G_2^\top) + \beta(1-\beta)(\widetilde{\psi}_0(k+1) + \widetilde{\psi}_1(k+1))$$
$$- \beta(1-\beta)(\psi_0(k+1)\psi_1(k+1)^\top + \psi_1(k+1)\psi_0(k+1)^\top))\bar{K}(k+1)^\top$$
$$- (\beta\bar{P}(k+1|k)C^\top + (1-\beta)A\bar{P}(k|k)C^\top)\bar{K}(k+1)^\top$$
$$- \bar{K}(k+1)(\beta C\bar{P}(k+1|k) + (1-\beta)C\bar{P}(k|k)A^\top), \tag{A.13}$$

where $\bar{P}(k+1|k)$, $\psi_i(k+1)$ and $\widetilde{\psi}_i(k+1)$, $\quad i = 0, 1$ are as defined in (3.28). To find the value of $\bar{K}(k+1)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(k+1|k+1)$ we differentiate the trace of the above expression with respect to the filter gain matrix $\bar{K}(k+1)$ and set the derivative to zero.

$$\frac{\partial tr\bar{P}(k+1|k+1)}{\partial\bar{K}(k+1)} = -2(\beta\bar{P}(k+1|k) + (1-\beta)A\bar{P}(k|k))C^\top + 2\bar{K}(k+1)[\beta(U_vU_v^\top +$$
$$C\bar{P}(k+1|k)C^\top + G_2q(k+1|k)G_2^\top) + (1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top + G_2q(k|k)G_2^\top) +$$
$$\beta(1-\beta)(\widetilde{\psi}_0(k+1) + \widetilde{\psi}_1(k+1)) - \beta(1-\beta)(\psi_0(k+1)\psi_1(k+1)^\top + \psi_1(k+1)\psi_0(k+1)^\top)]. \tag{A.14}$$

Setting $\frac{\partial tr \bar{P}(k+|k+1)}{\partial \bar{K}(k+1)} = 0$ leads the following expression for $\bar{K}(k+1)$:

$$\bar{K}(k+1) = (\beta \bar{P}(k+1|k) + (1-\beta)A\bar{P}(k|k))C^\top [\beta(U_v U_v^\top + C\bar{P}(k+1|k)C^\top + G_2 q(k+1|k)G_2^\top) + (1-\beta)(C\bar{P}(k|k)C^\top + U_v U_v^\top + G_2 q(k|k)G_2^\top) + \beta(1-\beta)(\widetilde{\psi}_0(k+1) + \widetilde{\psi}_1(k+1)) - \beta(1-\beta)(\psi_0(k+1)\psi_1(k+1)^\top + \psi_1(k+1)\psi_0(k+1)^\top)]^{-1}, \tag{A.15}$$

which is the required result. ∎

Proof of Theorem 3

This proof follows along the same lines as the proof of earlier theorems although it is notationally more involved. The approximated conditional mean of $\mathcal{X}(k+1)$, which provides the predictor, $\hat{\mathcal{X}}(k+1|k)$, is derived using (3.1):

$$\hat{\mathcal{X}}(k+1|k) = A\hat{\mathcal{X}}(k|k) + B. \tag{A.16}$$

In the case that the measurements are randomly delayed by N sampling times, the update equation for a linear filter using N step randomly delayed measurement is

$$\begin{aligned}\hat{\mathcal{X}}(k+1|k+1) &= \hat{\mathcal{X}}(k+1|k) + \bar{K}(k+1)(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k)) \\ &= A\hat{\mathcal{X}}(k|k) + B + \bar{K}(k+1)(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k)),\end{aligned} \tag{A.17}$$

where $\bar{K}(k+1)$ is derived by minimizing the trace of the estimation error covariance matrix and $\mathcal{Z}(k+1)$ is as defined in (3.29).

Next, note that

$$\bar{P}(k+1|k+1) = \mathbb{E}[(\mathcal{X}(k+1) - \hat{\mathcal{X}}(k+1|k+1))(\star)^\top]$$
$$= \mathbb{E}[(A\mathcal{X}(k) + B + U_w \mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k) - (A\hat{\mathcal{X}}(k|k) + B + \bar{K}(k+1)(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k))))(\star)^\top]$$

$$= \mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(\star)^\top A^\top] + U_w U_w^\top + G_1 q(k|k) G_1^\top + \bar{K}(k+1)\mathbb{E}[(\mathcal{Z}(k+1)$$
$$- \hat{\mathcal{Z}}(k+1|k))(\star)^\top]\bar{K}(k+1)^\top - \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w \mathcal{W}(k)$$
$$+ G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k))^\top \bar{K}(k+1)^\top - \bar{K}(k+1)\mathbb{E}[(\mathcal{Z}(k+1)$$
$$- \hat{\mathcal{Z}}(k+1|k))(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w \mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))^\top]. \quad \text{(A.18)}$$

Using Remark 4, we can write our one step ahead prediction of $\mathcal{Z}(k+1)$ in terms of the past predictions of $\hat{\mathcal{Y}}(k+1-i|k)$ as

$$\begin{aligned}\hat{\mathcal{Z}}(k+1|k) &= \beta\hat{\mathcal{Y}}(k+1|k) + (1-\beta)\beta\hat{\mathcal{Y}}(k|k) + (1-\beta)^2\beta\hat{\mathcal{Y}}(k-1|k) \\ &\quad + \cdots + (1-\beta)^{N-1}\beta\hat{\mathcal{Y}}(k-N+1|k) + (1-\beta)^N\hat{\mathcal{Y}}(k-N+1|k) \\ &= \beta\sum_{i=0}^{N-1}(1-\beta)^i\hat{\mathcal{Y}}(k+1-i|k) + (1-\beta)^N\hat{\mathcal{Y}}(k-N+1|k). \quad \text{(A.19)}\end{aligned}$$

Combining and re-arranging,

$$\begin{aligned}&(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k)) \\ &= \underbrace{\sum_{i=0}^{N-1}(\prod_{j=0}^{i} p_{k+1}^j)(1-p_{k+1}^{i+1})(\mathcal{Y}(k+1-i) - \hat{\mathcal{Y}}(k+1-i|k))}_{T_1} \\ &+ \underbrace{(\prod_{i=0}^{N} p_{k+1}^i)(\mathcal{Y}(k-N+1) - \hat{\mathcal{Y}}(k-N+1|k))}_{T_2} \\ &+ \underbrace{\sum_{i=0}^{N-1}\left((\prod_{j=0}^{i} p_{k+1}^j)(1-p_{k+1}^{i+1}) - \beta(1-\beta)^i\right)\hat{\mathcal{Y}}(k+1-i|k)}_{T_3} \\ &+ \underbrace{\left((\prod_{i=0}^{N} p_{k+1}^i) - (1-\beta)^N\right)\hat{\mathcal{Y}}(k-N+1|k)}_{T_4}, \quad \text{(A.20)}\end{aligned}$$

Next, using the notation defined in (A.20) and using the fact that p_{k+1}^i are i.i.d., we can easily show that

$$\hat{\chi}(k) := \mathbb{E}[(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k))(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k))^\top] =$$

$$\mathbb{E}(\sum_{i,j=1}^{4} T_i T_j^\top) = \mathbb{E}(\sum_{i=1}^{4} T_i T_i^\top) + \underbrace{\mathbb{E}(T_1T_2^\top + T_2T_1^\top)}_{\text{equals zero}} + \underbrace{\mathbb{E}(\sum_{j=3}^{4} T_1T_j^\top + T_jT_1^\top)}_{\text{equals zero}}$$

$$+ \underbrace{\mathbb{E}(\sum_{j=3}^{4} T_2T_j^\top + T_jT_2^\top)}_{\text{equals zero}} + \mathbb{E}(T_3T_4^\top + T_4T_3^\top). \tag{A.21}$$

where $\mathbb{E}[T_1T_1^\top] = \sum_{i=0}^{N-1} \hat{p}_k^i \mathbb{E}[(\mathcal{Y}(k+1-i) - \hat{\mathcal{Y}}(k+1-i|k))(\star)^\top]$

$$= \sum_{i=0}^{N-1} \mathbb{E}[(\prod_{j=0}^{i} p_{k+1}^j (1-p_{k+1}^{i+1}))^2]\mathbb{E}[(C\mathcal{X}(k+1-i) + D + U_v\mathcal{V}(k+1-i)$$
$$+ G_2 \operatorname{diag}(\mathcal{X}(k+1-i))\mathcal{S}_2(k+1-i) - (C\hat{\mathcal{X}}(k+1-i|k) + D))(\star)^\top]$$
$$= \beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)) + \beta(1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top$$
$$+ G_2q(k|k)G_2^\top)) + \sum_{i=2}^{N-1} \beta(1-\beta)^i \mathbb{E}[(C(A(A^{-i}(\mathcal{X}(k) - B - U_w\mathcal{W}(k-1) -$$
$$G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1) - (\hat{\mathcal{X}}(k|k) - B)) + \sum_{j=1}^{i-1} A^{-(i-j)}(-B - U_w\mathcal{W}(k-(j+1))$$
$$- G_1 \operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1)) + B)) + U_w\mathcal{W}(k-i) + G_1 \operatorname{diag}(\mathcal{X}(k-i))\mathcal{S}_1(k-i))$$
$$+ U_v\mathcal{V}(k+1-i) + G_2 \operatorname{diag}(\mathcal{X}(k+1-i))\mathcal{S}_2(k+1-i)))(\star)^\top)]$$
$$= \beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)) + \beta(1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top +$$
$$G_2q(k|k)G_2^\top)) + \sum_{i=2}^{N-1} \beta(1-\beta)^i \mathbb{E}[(C(A(A^{-i}(\mathcal{X}(k) - U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) +$$
$$\sum_{j=1}^{i-2} A^{-(i-j)}(-U_w\mathcal{W}(k-(j+1)) - G_1 \operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1)))) + U_v\mathcal{V}(k+1-i)$$

$$
\begin{aligned}
&+ G_2 \operatorname{diag}(\mathcal{X}(k+1-i))\mathcal{S}_2(k+1-i) - (CA^{-i+1}\hat{\mathcal{X}}(k|k)))(\star)^\top)] \\
&= \beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)) + \beta(1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top + \\
&G_2q(k|k)G_2^\top)) + \sum_{i=2}^{N-1}\beta(1-\beta)^i(C(A^{-i+1}(\bar{P}(k|k) + U_wU_w^\top + G_1q(k-1|k)G_1^\top)(A^{-i+1})^\top \\
&+ \sum_{j=1}^{i-2}A^{-(i-j)+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top)C^\top + U_vU_v^\top + G_2q(k+1-i|k)G_2^\top \\
&+ \mathbb{E}[(CA^{-i+1}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))(CA^{-i+1}(-U_w\mathcal{W}(k-1) - G_1\operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) + \\
&\sum_{j=1}^{i-2}A^{-(i-j)+1}(-U_w\mathcal{W}(k-(j+1)) - G_1\operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1))))^\top] \\
&+ \mathbb{E}[(CA^{-i+1}(-U_w\mathcal{W}(k-1) - G_1\operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) + \sum_{j=1}^{i-2}A^{-(i-j)+1}(-U_w\mathcal{W}(k-(j+1)) \\
&- G_1\operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1))))(CA^{-i+1}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))^\top]) \\
&= \beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)) + \beta(1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top + \\
&G_2q(k|k)G_2^\top)) + \sum_{i=2}^{N-1}\beta(1-\beta)^i(C(A^{-i+1}(\bar{P}(k|k) + U_wU_w^\top + G_1q(k-1|k)G_1^\top)(A^{-i+1})^\top + \\
&\sum_{j=1}^{i-2}A^{-(i-j)+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top)C^\top + U_vU_v^\top + G_2q(k+1-i|k)G_2^\top \\
&+ \mathbb{E}[(CA^{-i+1}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))(CA^{-i+1}(-U_w\mathcal{W}(k-1) - G_1\operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)))^\top] + \\
&\mathbb{E}[(CA^{-i+1}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))(\sum_{j=1}^{i-2}A^{-(i-j)+1}(-U_w\mathcal{W}(k-(j+1)) - \\
&G_1\operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1))))^\top]) + \mathbb{E}[(CA^{-i+1}(-U_w\mathcal{W}(k-1) - \\
&G_1\operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)))(CA^{-i+1}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))^\top] + \\
&\mathbb{E}[(\sum_{j=1}^{i-2}A^{-(i-j)+1}(-U_w\mathcal{W}(k-(j+1)) - G_1\operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1)))) \\
&(CA^{-i+1}(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))^\top]) \\
&= \beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)) + \beta(1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top + \\
&G_2q(k|k)G_2^\top)) + \sum_{i=2}^{N-1}\beta(1-\beta)^i(C(A^{-i+1}(\bar{P}(k|k) + U_wU_w^\top + G_1q(k-1|k)G_1^\top)(A^{-i+1})^\top \\
&+ \sum_{j=1}^{i-2}A^{-(i-j)+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top)C^\top + U_vU_v^\top + G_2q(k+1-i|k)G_2^\top \\
&- CA^{-i+1}(U_wU_w^\top + G_1q(k-1|k)G_1^\top)(A^{-i+1})^\top C^\top - \\
&(CA^{-i+1}\sum_{j=1}^{i-2}A^j(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top))
\end{aligned}
$$

$$- CA^{-i+1}(U_wU_w^\top + G_1q(k-1|k)G_1^\top)(A^{-i+1})^\top C^\top -$$

$$(C\sum_{j=1}^{i-2} A^{-(i-j)+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top))$$

$$= \beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)) + \beta(1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top +$$

$$G_2q(k|k)G_2^\top)) + \sum_{i=2}^{N-1}\beta(1-\beta)^i(C(A^{-i+1}(\bar{P}(k|k))(A^{-i+1})^\top C^\top + U_vU_v^\top + G_2q(k+1-i|k)G_2^\top$$

$$- CA^{-i+1}(U_wU_w^\top + G_1q(k-1|k)G_1^\top)(A^{-i+1})^\top C^\top -$$

$$(C\sum_{j=1}^{i-2} A^{-(i-j)+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top))$$

$$= \beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)) + \beta(1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top +$$

$$G_2q(k|k)G_2^\top)) + \sum_{i=2}^{N-1}\beta(1-\beta)^i(C(A^{-i+1}(\bar{P}(k|k))(A^{-i+1})^\top C^\top + U_vU_v^\top + G_2q(k+1-i|k)G_2^\top$$

$$(C\sum_{j=0}^{i-2} A^{-(i-j)+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top)) \qquad \text{(A.22)}$$

$$\mathbb{E}[T_2T_2^\top] = \hat{p}_k^N \mathbb{E}[(\mathcal{Y}(k-N+1) - \hat{\mathcal{Y}}(k-N+1|k))(\star)^\top]$$

$$= (1-\beta)^N \mathbb{E}[C\mathcal{X}(k+1-N) + D + U_v\mathcal{V}(k+1-N)$$

$$+ G_2 \operatorname{diag}(\mathcal{X}(k+1-N))\mathcal{S}_2(k+1-N) - (C\hat{\mathcal{X}}(k+1-N|k) + D))(\star)^\top]$$

$$= (1-\beta)^N \mathbb{E}[C(A^{-N}(\mathcal{X}(k+1) - B - U_w\mathcal{W}(k) - G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k)) +$$

$$\sum_{i=1}^{N-1} A^{-(N-i)}(-B - U_w\mathcal{W}(k-i) - G_1 \operatorname{diag}(\mathcal{X}(k-i)))\mathcal{S}_1(k-i))) + D + U_v\mathcal{V}(k+1-N)$$

$$+ G_2 \operatorname{diag}(\mathcal{X}(k+1-N))\mathcal{S}_2(k+1-N) - (CA^{-N}\hat{\mathcal{X}}(k+1|k) - B - B + D))(\star)^\top]$$

$$= (1-\beta)^N (C(A^{-N}(\bar{P}(k+1|k) + U_wU_w^\top + G_1q(k|k)G_1^\top)(A^{-N})^\top +$$

$$\sum_{i=1}^{N-1} A^{-(N-i)}(U_wU_w^\top + G_1q(k-i|k)G_1^\top)(A^{-(N-i)})^\top)C^\top + U_vU_v^\top + G_2q(k+1-N|k)G_2^\top +$$

$$\mathbb{E}[(CA^{-N}\mathcal{X}(k+1))(CA^{-N}(-U_w\mathcal{W}(k) - G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k)))^\top] +$$

$$\mathbb{E}[(CA^{-N}\mathcal{X}(k+1))(C\sum_{i=1}^{N-1}A^{-(N-i)}(-U_w\mathcal{W}(k-i)-G_1\,\mathrm{diag}(\mathcal{X}(k-i)))\mathcal{S}_1(k-i))^\top]+$$
$$\mathbb{E}[(CA^{-N}(-U_w\mathcal{W}(k)-G_1\,\mathrm{diag}(\mathcal{X}(k))\mathcal{S}_1(k)))(CA^{-N}\mathcal{X}(k+1))^\top]+$$
$$\mathbb{E}[(C\sum_{i=1}^{N-1}A^{-(N-i)}(-U_w\mathcal{W}(k-i)-G_1\,\mathrm{diag}(\mathcal{X}(k-i)))\mathcal{S}_1(k-i))(CA^{-N}\mathcal{X}(k+1))^\top])$$
$$=(1-\beta)^N(C(A^{-N}(\bar{P}(k+1|k)+U_wU_w^\top+G_1q(k|k)G_1^\top)(A^{-N})^\top+\sum_{i=1}^{N-1}A^{-(N-i)}(U_wU_w^\top$$
$$+G_1q(k-i|k)G_1^\top)(A^{-(N-i)})^\top)C^\top+U_vU_v^\top+G_2q(k+1-N|k)G_2^\top-$$
$$CA^{-N}(U_wU_w^\top+G_1q(k|k)G_1^\top)(A^{-N})^\top C^\top-CA^{-N}\sum_{i=1}^{N-1}A^{i}(U_wU_w^\top$$
$$+G_1q(k-i|k)G_1^\top)(A^{-(N-i)})^\top C^\top-CA^{-N}(U_wU_w^\top+G_1q(k|k)G_1^\top)(A^{-N})^\top C^\top$$
$$-C\sum_{i=1}^{N-1}A^{-(N-i)}(U_wU_w^\top+G_1q(k-i|k)G_1^\top)(A^{-(N-i)})^\top C^\top)$$
$$=(1-\beta)^N(CA^{-N}\bar{P}(k+1|k)(A^{-N})^\top C^\top+U_vU_v^\top+G_2q(k+1-N|k)G_2^\top$$
$$-CA^{-N}(U_wU_w^\top+G_1q(k|k)G_1^\top)(A^{-N})^\top C^\top$$
$$-C\sum_{i=1}^{N-1}A^{-(N-i)}(U_wU_w^\top+G_1q(k-i|k)G_1^\top)(A^{-(N-i)})^\top C^\top)$$
$$=(1-\beta)^N(CA^{-N}\bar{P}(k+1|k)(A^{-N})^\top C^\top+U_vU_v^\top+G_2q(k+1-N|k)G_2^\top$$
$$-C\sum_{i=0}^{N-1}A^{-(N-i)}(U_wU_w^\top+G_1q(k-i|k)G_1^\top)(A^{-(N-i)})^\top C^\top) \tag{A.23}$$

$$\mathbb{E}[T_3T_3^\top]=\sum_{i=0}^{N-1}\widetilde{\hat{p}}_k^i\mathbb{E}[\hat{\mathcal{Y}}(k+1-i|k))(\star)^\top]=\sum_{i=0}^{N-1}(\beta(1-\beta)^i-\beta^2(1-\beta)^{2i})\widetilde{\psi}_i(k+1), \tag{A.24}$$

$$\mathbb{E}[T_4T_4^\top]=\widetilde{\hat{p}}_k^N\mathbb{E}[\hat{\mathcal{Y}}(k-N+1|k)(\star)^\top]\quad=((1-\beta)^N-(1-\beta)^{2N})\widetilde{\psi}_N(k+1), \tag{A.25}$$

$$\mathbb{E}(T_3T_4^\top) = -\beta(1-\beta)^N \sum_{i=0}^{N-1}(1-\beta)^i\psi_i(k+1)\psi_N(k+1)^\top, \tag{A.26}$$

$$\mathbb{E}(T_4T_3^\top) = -\beta(1-\beta)^N \psi_N(k+1)\sum_{i=0}^{N-1}(1-\beta)^i\psi_i(k+1)^\top, \tag{A.27}$$

where

$$\begin{aligned}
&\bar{P}(k+1|k) = A\bar{P}(k|k)A^\top + U_wU_w^\top + G_1q(k|k)G_1^\top,\\
&\hat{\mathcal{Y}}(k+1-i|k) = C\hat{\mathcal{X}}(k+1-i|k) + D =: \psi_i(k+1),\\
&\widetilde{\psi}_i(k+1) = \psi_i(k+1)\psi_i(k+1)^\top,
\end{aligned} \tag{A.28}$$

Substituting (A.22) - (A.27) at (A.21) we have

$$\begin{aligned}
&\hat{\chi}(k) := \mathbb{E}(\mathcal{Z}(k+1) - \hat{\mathcal{Z}}(k+1|k))(\star)^\top\\
&= \beta(C\bar{P}(k+1|k)C^\top + U_vU_v^\top + G_2q(k+1|k)G_2^\top)) + \beta(1-\beta)(C\bar{P}(k|k)C^\top + U_vU_v^\top +\\
&G_2q(k|k)G_2^\top)) + \sum_{i=2}^{N-1}\beta(1-\beta)^i(C(A^{-i+1}(\bar{P}(k|k))(A^{-i+1})^\top C^\top + U_vU_v^\top + G_2q(k+1-i|k)G_2^\top\\
&-(C\sum_{j=0}^{i-2}A^{-(i-j)+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top))\\
&+(1-\beta)^N(CA^{-N}\bar{P}(k+1|k)(A^{-N})^\top C^\top + U_vU_v^\top + G_2q(k+1-N|k)G_2^\top\\
&-C\sum_{i=0}^{N-1}A^{-(N-i)}(U_wU_w^\top + G_1q(k-i|k)G_1^\top)(A^{-(N-i)})^\top C^\top)\\
&+\sum_{i=0}^{N-1}(\beta(1-\beta)^i - \beta^2(1-\beta)^{2i})\widetilde{\psi}_i(k+1) + ((1-\beta)^N - (1-\beta)^{2N})\widetilde{\psi}_N(k+1)\\
&-\beta(1-\beta)^N\sum_{i=0}^{N-1}(1-\beta)^i\psi_i(k+1)\psi_N(k+1)^\top - \beta(1-\beta)^N\psi_N(k+1)\sum_{i=0}^{N-1}(1-\beta)^i\psi_i(k+1)^\top
\end{aligned} \tag{A.29}$$

We need to evaluate the following expectation to get an expression for $\bar{P}(k+$

$1|k+1)$:

$$\mathbb{E}[(A(\mathcal{X}(k)-\hat{\mathcal{X}}(k|k))+U_w\mathcal{W}(k)+G_1\operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(\sum_{i=0}^{N-1}(\prod_{j=0}^{i}p_{k+1}^{j})(1-p_{k+1}^{i+1})(\mathcal{Y}(k+1-i)$$

$$-\hat{\mathcal{Y}}(k+1-i|k))+\prod_{i=0}^{N}(p_{k+1}^{i})(\mathcal{Y}(k-N+1)-\hat{\mathcal{Y}}(k-N+1|k))+$$

$$\sum_{i=0}^{N-1}\left(\prod_{j=0}^{i}p_{k+1}^{j}(1-p_{k+1}^{i+1})-\beta(1-\beta)^{i}\right)\psi_i(k+1)+\left((\prod_{i=0}^{N}p_{k+1}^{i})-(1-\beta)^{N}\right)\psi_N(k+1))^{\top}]$$

$$=\sum_{i=0}^{N-1}\beta(1-\beta)^{i}\mathbb{E}[(A(\mathcal{X}(k)-\hat{\mathcal{X}}(k|k))+U_w\mathcal{W}(k)+G_1\operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))$$

$$(C\mathcal{X}(k+1-i)+U_v\mathcal{V}(k+1-i)+G_2\operatorname{diag}(\mathcal{X}(k+1-i))\mathcal{S}_2(k+1-i)-CA\hat{\mathcal{X}}(k-i|k))^{\top}]$$

$$+\mathbb{E}[(A(\mathcal{X}(k)-\hat{\mathcal{X}}(k|k))+U_w\mathcal{W}(k)+G_1\operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))\times$$

$$(\prod_{i=0}^{N}p_{k+1}^{i}(\mathcal{Y}(k-N+1)-\hat{\mathcal{Y}}(k-N+1|k)))^{\top}]+\sum_{i=0}^{N-1}\mathbb{E}[(A(\mathcal{X}(k)-\hat{\mathcal{X}}(k|k))+U_w\mathcal{W}(k)+$$

$$G_1\operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))\underbrace{(\left(\prod_{j=0}^{i}p_{k+1}^{j}(1-p_{k+1}^{i+1})-\beta(1-\beta)^{i}\right)\psi_i(k+1))^{\top}]}_{\text{equals zero}}+\mathbb{E}[(A(\mathcal{X}(k)-\hat{\mathcal{X}}(k|k))$$

$$+U_w\mathcal{W}(k)+G_1\operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))\underbrace{(\left((\prod_{i=0}^{N}p_{k+1}^{i})-(1-\beta)^{N}\right)\psi_N(k+1)^{\top})}_{\text{equals zero}}$$

$$=\beta(\bar{P}(k+1|k)C^{\top})+\beta(1-\beta)A\bar{P}(k|k)C^{\top}+\sum_{i=2}^{N-1}\beta(1-\beta)^{i}(\mathbb{E}[(A(\mathcal{X}(k)-\hat{\mathcal{X}}(k|k))+$$

$$U_w\mathcal{W}(k)+G_1\operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(C\mathcal{X}(k+1-i)+U_v\mathcal{V}(k+1-i)+$$

$$G_2\operatorname{diag}(\mathcal{X}(k+1-i))\mathcal{S}_2(k+1-i)-C\hat{\mathcal{X}}(k+1-i|k))^{\top}])+$$

$$(1-\beta)^{N}(\mathbb{E}[(A(\mathcal{X}(k)-\hat{\mathcal{X}}(k|k))+U_w\mathcal{W}(k)+G_1\operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(C(\mathcal{X}(k+1-N)+$$

$$D+U_v\mathcal{V}(k+1-N)+G_2\operatorname{diag}(\mathcal{X}(k+1-N))\mathcal{S}_2(k+1-N)-\mathcal{X}(k+1-N|k))^{\top}]$$

$$= \beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1} \beta(1-\beta)^i(\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))\times$$

$$(C\mathcal{X}(k+1-i) - C\hat{\mathcal{X}}(k+1-i|k))^\top]) + (1-\beta)^N(\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))\times$$

$$(C(\mathcal{X}(k+1-N) - \mathcal{X}(k+1-N|k))^\top]$$

To find $\mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(C(\mathcal{X}(k+1-i) - \hat{\mathcal{X}}(k+1-i|k)))^\top]$ and $\mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(C(\mathcal{X}(k+1-N) - \hat{\mathcal{X}}(k+1-N|k)))^\top]$, we will use lemma 2:

Hence $\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1 \operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))(\sum_{i=0}^{N-1}(\prod_{j=0}^{i} p_{k+1}^j)(1-p_{k+1}^{i+1})$

$$(\mathcal{Y}(k+1-i) - \hat{\mathcal{Y}}(k+1-i|k)) + \prod_{i=0}^{N}(p_{k+1}^i)(\mathcal{Y}(k-N+1) - \hat{\mathcal{Y}}(k-N+1|k)) +$$

$$\sum_{i=0}^{N-1}\left(\prod_{j=0}^{i} p_{k+1}^j(1-p_{k+1}^{i+1}) - \beta(1-\beta)^i\right)\psi_i(k+1) + \left((\prod_{i=0}^{N} p_{k+1}^i) - (1-\beta)^N\right)\psi_N(k+1))^\top]$$

$$= \beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1} \beta(1-\beta)^i(\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))(CA(A^{-i}(\mathcal{X}(k)$$

$$- B - U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) + \sum_{j=1}^{i-1} A^{-(i-j)}(-B - U_w\mathcal{W}(k-(j+1))$$

$$- G_1 \operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1)))) + U_w\mathcal{W}(k-i) + G_1 \operatorname{diag}(\mathcal{X}(k-i))\mathcal{S}_1(k-i))$$

$$- (CAA^{-i}\hat{\mathcal{X}}(k|k) - B - B))^\top]) + (1-\beta)^N(\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))(CA(A^{-N}(\mathcal{X}(k) - B -$$

$$U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) + \sum_{i=1}^{N-1} A^{-(N-i)}(-B - U_w\mathcal{W}(k-(i+1))$$

$$- G_1 \operatorname{diag}(\mathcal{X}(k-(i+1)))\mathcal{S}_1(k-(i+1))) + U_w\mathcal{W}(k-N)$$

$$+ G_1 \operatorname{diag}(\mathcal{X}(k-N))\mathcal{S}_1(k-N)) - (CAA^{-N}\mathcal{X}(k|k) - B - B))^\top]$$

$$= \beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1} \beta(1-\beta)^i(\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))$$

$$(C(A^{-i+1}(\mathcal{X}(k) - U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) +$$

$$\sum_{j=1}^{i-2} A^{-(i-j)+1}(-U_w\mathcal{W}(k-(j+1)) - G_1 \operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1))))$$
$$- CAA^{-i}\hat{\mathcal{X}}(k|k))^\top]) + (1-\beta)^N(\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))$$
$$(CA(A^{-N}(\mathcal{X}(k) - U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) + \sum_{i=1}^{N-2} A^{-(N-i)+1}(-$$
$$U_w\mathcal{W}(k-(i+1)) - G_1 \operatorname{diag}(\mathcal{X}(k-(i+1)))\mathcal{S}_1(k-(i+1))) - CAA^{-N}\mathcal{X}(k|k))^\top]$$
$$= \beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1} \beta(1-\beta)^i(A\bar{P}(k|k)(A^{-i+1})^\top C^\top + \mathbb{E}[A(\mathcal{X}(k)$$
$$- \hat{\mathcal{X}}(k|k))(CA^{-i+1}(-U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)))^\top] + \mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))\times$$
$$(\sum_{j=1}^{i-2} A^{-(i-j)+1}(-U_w\mathcal{W}(k-(j+1)) - G_1 \operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1))))^\top]$$
$$+ (1-\beta)^N(A\bar{P}(k|k)(A^{-N+1})^\top C^\top + \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))(CA^{-N+1}(-U_w\mathcal{W}(k-1)$$
$$- G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)))^\top] + \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))(\sum_{i=1}^{N-2} A^{-(N-i)+1}\times$$
$$(-U_w\mathcal{W}(k-(i+1)) - G_1 \operatorname{diag}(\mathcal{X}(k-(i+1)))\mathcal{S}_1(k-(i+1))))^\top]$$
$$= \beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1} \beta(1-\beta)^i(A\bar{P}(k|k)(A^{-i+1})^\top C^\top +$$
$$\mathbb{E}[A(A\mathcal{X}(k-1) + U_w\mathcal{W}(k-1) + G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) - A\hat{\mathcal{X}}(k-1|k))$$
$$\times (CA^{-i+1}(-U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)))^\top]+$$
$$\mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(C\sum_{j=1}^{i-2} A^{-(i-j)+1}(-U_w\mathcal{W}(k-(j+1)) -$$
$$G_1 \operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1))))^\top] + (1-\beta)^N(A\bar{P}(k|k)(A^{-N+1})^\top C^\top +$$
$$\mathbb{E}[(A(A\mathcal{X}(k-1) + U_w\mathcal{W}(k-1) + G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)) - A\hat{\mathcal{X}}(k-1|k))$$
$$(CA^{-N+1}(-U_w\mathcal{W}(k-1) - G_1 \operatorname{diag}(\mathcal{X}(k-1))\mathcal{S}_1(k-1)))^\top]+$$
$$\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)))(C\sum_{i=1}^{N-2} A^{-(N-i)+1}(-B - U_w\mathcal{W}(k-(i+1))$$
$$- G_1 \operatorname{diag}(\mathcal{X}(k-(i+1)))\mathcal{S}_1(k-(i+1))))^\top]$$

$$
\begin{aligned}
&= \beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1}\beta(1-\beta)^i(A\bar{P}(k|k)A^{-i+1}C^\top - A(U_wU_w^\top + \\
&G_1q(k-1|k)G_1^\top)(A^{-i+1})^\top C^\top + \mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(C\sum_{j=1}^{i-2}A^{-(i-j)+1}(-U_w\mathcal{W}(k-(j+1)) \\
&- G_1 \operatorname{diag}(\mathcal{X}(k-(j+1)))\mathcal{S}_1(k-(j+1))))^\top] + (1-\beta)^N(A\bar{P}(k|k)(A^{-N+1})^\top C^\top - \\
&A(U_wU_w^\top + G_1q(k-1|k)G_1^\top)(A^{-N+1})^\top C^\top + \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))) \\
&(C\sum_{i=1}^{N-2}A^{-(N-i)+1}(-B - U_w\mathcal{W}(k-(i+1)) - G_1 \operatorname{diag}(\mathcal{X}(k-(i+1)))\mathcal{S}_1(k-(i+1))))^\top] \\
&= \beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1}\beta(1-\beta)^i(A\bar{P}(k|k)A^{-i+1}C^\top - A(U_wU_w^\top + \\
&G_1q(k-1|k)G_1^\top)(A^{-i+1})^\top C^\top - \sum_{j=1}^{i-2}A^{j+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top + \\
&(1-\beta)^N(A\bar{P}(k|k)(A^{-N+1})^\top C^\top - A(U_wU_w^\top + G_1q(k-1|k)G_1^\top)(A^{-N+1})^\top C^\top \\
&- \sum_{i=1}^{N-2}A^{i+1}(U_wU_w^\top + G_1q(k-(i+1)|k)G_1^\top)(A^{-(N-i)+1})^\top C^\top \\
&= \beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=2}^{N-1}\beta(1-\beta)^i(A\bar{P}(k|k)(A^{-i+1})^\top C^\top - \\
&\sum_{j=0}^{i-2}A^{j+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top + (1-\beta)^N(A\bar{P}(k|k)(A^{-N+1})^\top C^\top \\
&- \sum_{i=0}^{N-2}A^{i+1}(U_wU_w^\top + G_1q(k-(i+1)|k)G^\top)(A^{-(N-i)+1})^\top C^\top \qquad \text{(A.30)}
\end{aligned}
$$

where, as before,

$$\hat{\mathcal{Y}}(k+1-i|k) = \psi_i(k+1). \qquad \text{(A.31)}$$

Similarly,

$$\mathbb{E}[(\sum_{i=0}^{N-1}(\prod_{j=0}^{i}p_{k+1}^{j})(1-p_{k+1}^{i+1})(\mathcal{Y}(k+1-i)-\hat{\mathcal{Y}}(k+1-i|k))+\prod_{i=0}^{N}(p_{k+1}^{i})(\mathcal{Y}(k-N+1)-$$
$$\hat{\mathcal{Y}}(k-N+1|k))+\sum_{i=0}^{N-1}\left(\prod_{j=0}^{i}p_{k+1}^{j}(1-p_{k+1}^{i+1})-\beta(1-\beta)^{i}\right)\psi_i(k+1)+$$
$$\left((\prod_{i=0}^{N}p_{k+1}^{i})-(1-\beta)^{N}\right)\psi_N(k+1))(A(\mathcal{X}(k)-\hat{\mathcal{X}}(k|k))+U_w\mathcal{W}(k)+G_1\operatorname{diag}(\mathcal{X}(k))\mathcal{S}_1(k))^{\top}]$$
$$=\beta(C\bar{P}(k+1|k))+\beta(1-\beta)C\bar{P}(k|k)A^{\top}+\sum_{i=2}^{N-1}\beta(1-\beta)^{i}(CA^{-i+1}\bar{P}(k|k)A^{\top}-$$
$$\sum_{j=0}^{i-2}CA^{-(i-j)+1}(U_wU_w^{\top}+G_1q(k-(j+1)|k)G_1^{\top})(A^{j+1})^{\top}+(1-\beta)^{N}CA^{-N+1}(\bar{P}(k|k)A^{\top}$$
$$-\sum_{i=0}^{N-2}CA^{-(N-i)+1}(U_wU_w^{\top}+G_1q(k-(i+1)|k)G_1^{\top})(A^{i+1})^{\top}. \tag{A.32}$$

After substituting (A.29)- (A.32) in (A.18), we arrive at the following expression for the estimation error covariance matrix:

$$\bar{P}(k+1|k+1)=\bar{P}(k+1|k)+\bar{K}(k+1)\hat{\chi}(k+1)\bar{K}(k+1)^{\top}-(\beta(\bar{P}(k+1|k)C^{\top})+$$
$$\beta(1-\beta)A\bar{P}(k|k)C^{\top}+\sum_{i=2}^{N-1}\beta(1-\beta)^{i}(A\bar{P}(k|k)(A^{-i+1})^{\top}C^{\top}-\sum_{j=0}^{i-2}A^{j+1}(U_wU_w^{\top}$$
$$+G_1q(k-(j+1)|k)G_1^{\top})(A^{-(i-j)+1})^{\top}C^{\top})+(1-\beta)^{N}(A\bar{P}(k|k)(A^{-N+1})^{\top}C^{\top}-$$
$$\sum_{i=0}^{N-2}A^{i+1}(U_wU_w^{\top}+G_1q(k-(i+1))G_1^{\top})(A^{-(N-i)+1})^{\top}C^{\top}))\bar{K}(k+1)^{\top}$$
$$-\bar{K}(k+1)(\beta(C\bar{P}(k+1|k))+\beta(1-\beta)C\bar{P}(k|k)A^{\top}+\sum_{i=1}^{N-1}\beta(1-\beta)^{i}(CA^{-i+1}\bar{P}(k|k)A^{\top}-$$
$$\sum_{j=0}^{i-2}CA^{-(i-j)+1}(U_wU_w^{\top}+G_1q(k-(j+1)|k)G_1^{\top})(A^{j+1})^{\top}+(1-\beta)^{N}CA^{-N+1}(\bar{P}(k|k)A^{\top}$$
$$-\sum_{i=0}^{N-2}CA^{-(N-i)+1}(U_wU_w^{\top}+G_1q(k-(i+1)|k)G_1^{\top})(A^{i+1})^{\top}), \tag{A.33}$$

To find the value of $\bar{K}(k+1)$ that minimizes the trace of the covariance $\bar{P}(k+1|k+1)$ we differentiate the trace of the above expression with respect to the filter gain matrix $\bar{K}(k+1)$ and set the derivative to zero.

$$\frac{\partial tr\bar{P}(k+|k+1)}{\partial\bar{K}(k+1)} = -2(\beta(\bar{P}(k+1|k)C^\top) + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=1}^{N-1}\beta(1-\beta)^i(A\bar{P}(k|k)\times$$
$$(A^{-i+1})^\top C^\top - \sum_{j=0}^{i-2}A^{j+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top)$$
$$+(1-\beta)^N(A\bar{P}(k|k)(A^{-N+1})^\top C^\top - \sum_{i=0}^{N-2}A^{i+1}(U_wU_w^\top + G_1q(k-(i+1)|k)G_1^\top)\times$$
$$(A^{-(N-i)+1})^\top C^\top)) + 2\bar{K}(k+1)\hat{\chi}(k+1) \tag{A.34}$$

Setting $\frac{\partial tr\bar{P}(k+|k+1)}{\partial\bar{K}(k+1)} = 0$ leads the following expression for $\bar{K}(k+1)$:

$$\bar{K}(k+1) = (\beta\bar{P}(k+1|k)C^\top + \beta(1-\beta)A\bar{P}(k|k)C^\top + \sum_{i=1}^{N-1}\beta(1-\beta)^i(A\bar{P}(k|k)(A^{-i+1})^\top C^\top -$$
$$\sum_{j=0}^{i-2}A^{j+1}(U_wU_w^\top + G_1q(k-(j+1)|k)G_1^\top)(A^{-(i-j)+1})^\top C^\top) + (1-\beta)^N(A\bar{P}(k|k)(A^{-N+1})^\top C^\top$$
$$-\sum_{i=0}^{N-2}A^{i+1}(U_wU_w^\top + G_1q(k-(i+1)|k)G_1^\top)(A^{-(N-i)+1})^\top C^\top))\hat{\chi}(k+1)^{-1} \tag{A.35}$$

where $\bar{P}(k+1|k)$ is as defined in (A.28).
This concludes the proof of theorem 3. ∎

Proof of Theorem 4

This proof follows along the same lines as the proof of earlier theorems. The filtering estimates of the state covariance is obtained by combining the equations (4.2)-(4.18) as follows. The estimation error covariance matrix at time $k+1$ can be written as

$$
\begin{aligned}
&\bar{P}(t_k^{i+1}|t_{k+1}) = \mathbb{E}[(\mathcal{X}(t_k^{i+1}) - \mathcal{X}(t_k^{i+1}|t_{k+1}))(\star)^\top] \\
&= \mathbb{E}[\tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m\mathcal{W}(t_k^i) + \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^i) - \\
&(\tilde{A}\hat{\mathcal{X}}(t_k^i|t_k) + \tilde{B} + \bar{K}(t_{k+1}^i)(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k)))(\star)^\top] \\
&= \mathbb{E}[\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k))(\star)^\top] + \tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top + \\
&\bar{K}(t_{k+1}^i)(\mathbb{E}(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k))(\star)^\top)\bar{K}(t_{k+1}^i)^\top - \\
&(\tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m\mathcal{W}(t_k^i) + \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k) - B) \\
&(\bar{K}(t_{k+1}^i)(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k)))^\top - (\bar{K}(t_{k+1}^i)(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k))) \\
&(\tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m\mathcal{W}(t_k^i) + \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k) - B)^\top \qquad \text{(A.36)}
\end{aligned}
$$

Next, we need the following covariance term in evaluating $\bar{P}(t_k^{i+1}|t_k)$:

$$
\mathbb{E}[(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k))(\star)^\top)] = C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top \qquad \text{(A.37)}
$$

We also need to evaluate some cross covariance terms, whose expressions are derived next:

$$
\begin{aligned}
&\mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\mathcal{W}(t_k^i) + \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^i))(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k))^\top] \\
&= \mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\mathcal{W}(t_k^i) + \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^i)) \\
&(C(\mathcal{X}(t_k^m) - \hat{\mathcal{X}}(t_k^m|t_k)) + U_v\mathcal{V}(t_k^m))^\top] \\
&= \mathbb{E}[(\tilde{A}(\tilde{A}^{i-m}(\mathcal{X}(t_k^m) - \tilde{B} - \tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-1}) - \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))\tilde{\mathcal{S}}(t_k^{m-1}))) + \\
&\sum_{r=1}^{m-i-1} \tilde{A}^{-(m-i-r)}(-\tilde{B} - \tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-(r+1)}) - \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))\tilde{\mathcal{S}}(t_k^{m-(r+1)})) + B) \\
&- (\tilde{A}^{i-m+1}\mathcal{X}(t_k^m|t_k)) - \tilde{B}) + \tilde{U}_m\mathcal{W}(t_k^i) + \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^i)) \\
&(C(\mathcal{X}(t_k^m) - \hat{\mathcal{X}}(t_k^m|t_k)) + U_v\mathcal{V}(t_k^m)) \\
&= \mathbb{E}[(\tilde{A}(\tilde{A}^{i-m}(\mathcal{X}(t_k^m) - \tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-1}) - \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))\tilde{\mathcal{S}}(t_k^{m-1}))) + \\
&\sum_{r=1}^{m-i-2} \tilde{A}^{-(m-i-r)}(-\tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-(r+1)}) - \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))\tilde{\mathcal{S}}(t_k^{m-(r+1)}))
\end{aligned}
$$

$$- (\tilde{A}^{i-m+1}\mathcal{X}(t_k^m|t_k)))(C(\mathcal{X}(t_k^m) - \hat{\mathcal{X}}(t_k^m|t_k)) + U_v\mathcal{V}(t_k^m))$$
$$= \tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k) - \tilde{U}_m\tilde{U}_m^\top - \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))^\top]\tilde{G}^\top)C^\top$$
$$- \sum_{r=1}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))^\top]\tilde{G}^\top)(\tilde{A}^{(r)})^\top C^\top \tag{A.38}$$

Further,

$$\mathbb{E}[(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k))(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\mathcal{W}(t_k^{i+1}) + \tilde{G}\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^{i+1}))^\top]$$
$$= C(\bar{P}(t_k^m|t_k) - \tilde{U}_m\tilde{U}_m^\top - \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))^\top]\tilde{G}^\top)(\tilde{A}^{i-m+1})^\top$$
$$- \sum_{r=1}^{m-i-2} C(\tilde{A}^r)(\tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))^\top]\tilde{G}^\top)(\tilde{A}^{-(m-i-r)+1})^\top \tag{A.39}$$

Substituting (A.37), (A.38) and (A.39) in (A.36), we have

$$\bar{P}(t_k^{i+1}|t_k) = \tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top + \tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top$$
$$+ \bar{K}(t_{k+1}^i)(C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top)\bar{K}(t_{k+1}^i)^\top - (\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k) - \tilde{U}_m\tilde{U}_m^\top$$
$$- \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))^\top]\tilde{G}^\top)C^\top - \sum_{r=1}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_m\tilde{U}_m^\top +$$
$$\tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))^\top]\tilde{G}^\top)(\tilde{A}^{(r)})^\top C^\top)\bar{K}^\top(t_{k+1}^i) -$$
$$\bar{K}(t_{k+1}^i)(C(\bar{P}(t_k^m|t_k) - \tilde{U}_m\tilde{U}_m^\top - \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))^\top]\tilde{G}^\top)(\tilde{A}^{i-m+1})^\top$$
$$- \sum_{r=1}^{m-i-2} C(\tilde{A}^r)(\tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))^\top]\tilde{G}^\top)(\tilde{A}^{-(m-i-r)+1})^\top)$$

where $\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]$ is as in equation (9). To find the value of $\bar{K}(t_{k+1}^i)$ that minimizes the trace of the covariance $\bar{P}(t_k^{i+1}|t_k)$ we differentiate the trace of the above expression with respect to the filter gain matrix $\bar{K}(t_{k+1}^i)$ and set the derivative to zero.

$$\frac{\partial tr\bar{P}(t_k^{i+1}|t_k)}{\partial\bar{K}(t_{k+1}^i)} = -2(\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k) - \tilde{U}_m\tilde{U}_m^\top - \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))^\top]\tilde{G}^\top)C^\top$$
$$- \sum_{r=1}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))^\top]\tilde{G}^\top)(\tilde{A}^{(r)})^\top C^\top)$$
$$+ 2\bar{K}(t_{k+1}^i)(C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top) \tag{A.40}$$

Setting this partial derivative to zero leads the following expression for $\bar{K}(t_{k+1}^i)$:

$$\bar{K}(t_{k+1}^i) = (\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k) - \tilde{U}_m\tilde{U}_m^\top - \tilde{G}\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^{m-1}))^\top]\tilde{G}^\top)C^\top - \sum_{r=1}^{m-i-2}\tilde{A}^{-(m-i-r)+1}(\tilde{U}_m\tilde{U}_m^\top + \tilde{G}\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^{m-(r+1)}))^\top]\tilde{G}^\top)(\tilde{A}^{(r)})^\top C^\top) [C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top]^{-1}. \tag{A.41}$$

which is the required result. ∎

Proof of Theorem 5

The proof follows on the same lines as the proof of earlier theorems . The filtering estimates of the state covariance is obtained by combining the equations (4.12)-(4.18) as follows. The state covariance matrix at time t_{k+1} can be written as

$$\begin{aligned}
&\bar{P}(t_k^{i+1}|t_{k+1}) = \mathbb{E}[(\mathcal{X}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}))((\mathcal{X}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}))^\top] \\
&= \mathbb{E}[(\tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i)) - (\tilde{A}\hat{\mathcal{X}}(t_k^i|t_k) + \tilde{B} \\
&+ \bar{K}(t_{k+1}^i)(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k)))(\star)^\top] \\
&= \mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)))(\star)^\top] + \tilde{U}_m\tilde{U}_m^\top + \tilde{G}_1q(t_k^i)\tilde{G}_1^\top + \bar{K}(t_{k+1}^i)\mathbb{E}[(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))(\star)^\top]\times \\
&\bar{K}(t_{k+1}^i)^\top - \mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))^\top]\times \\
&\bar{K}(t_{k+1}^i)^\top - \bar{K}(t_{k+1}^i)\mathbb{E}[(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))^\top]
\end{aligned} \tag{A.42}$$

Using the independence of p_{jk} with each other and using $\mathbb{E}(p_{jk}^2) = 1 - \beta$, we have

$$\mathbb{E}\left(\prod_{i=0}^{j} p_{ik}(1 - p_{jk})\right)^2 = \beta(1-\beta)^j, \tag{A.43}$$

$$\mathbb{E}\left(\prod_{j=0}^{N} p_{jk}\right)^2 = (1-\beta)^N, \tag{A.44}$$

$$\mathbb{E}\left(\prod_{i=0}^{j} p_{ik}(1-p_{(j+1)k}) - \beta(1-\beta)^j\right)^2 = \beta(1-\beta)^j - \beta^2(1-\beta)^{2j}, \tag{A.45}$$

$$\mathbb{E}\left(\prod_{j=0}^{N-1} p_{jk} - (1-\beta)^N\right)^2 = (1-\beta)^N - (1-\beta)^{2N}. \tag{A.46}$$

Next, using the notation defined in (4.16) and using the fact that p_{jk} are i.i.d., we can easily show that

$$\begin{aligned}
&\mathbb{E}[(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))^\top] = \\
&\mathbb{E}(\sum_{i,j=1}^{4} T_i T_j^\top) = \mathbb{E}(\sum_{i=1}^{4} T_i T_i^\top) + \underbrace{\mathbb{E}(T_1 T_2^\top + T_2 T_1^\top)}_{\text{equals zero}} + \underbrace{\mathbb{E}(\sum_{j=3}^{4} T_1 T_j^\top + T_j T_1^\top)}_{\text{equals zero}} \\
&+ \underbrace{\mathbb{E}(\sum_{j=3}^{4} T_2 T_j^\top + T_j T_2^\top)}_{\text{equals zero}} + \mathbb{E}(T_3 T_4^\top + T_4 T_3^\top).
\end{aligned} \tag{A.47}$$

where

$$\begin{aligned}
&\mathbb{E}[T_1 T_1^\top] = \mathbb{E}[\sum_{j=0}^{N-1} (\prod_{i=0}^{j} p_{ik})(1-p_{(j+1)k})(\mathcal{Y}(t_k^{m-j}) - \hat{\mathcal{Y}}(t_k^{m-j}|t_k))(\star)^\top] \\
&= \sum_{j=0}^{N-1} \mathbb{E}[(\prod_{i=0}^{j} p_{ik}(1-p_{(j+1)k}))^2]\mathbb{E}[(\mathcal{Y}(t_k^{m-j}) - \hat{\mathcal{Y}}(t_k^{m-j}|t_k))(\star)^\top] \\
&= \sum_{j=0}^{N-1} \beta(1-\beta)^j \mathbb{E}[(C\mathcal{X}(t_k^{m-j}) + U_v\mathcal{V}(t_k^{m-j}) + G_2 \operatorname{diag}(\mathcal{X}(t_k^{m-j}))\mathcal{S}_2(t_k^{m-j}) - C\hat{\mathcal{X}}(t_k^{m-j}|t_k))(\star)^\top] \\
&= \beta(C\bar{P}(t_k^m|t_k)C^\top + U_v U_v^\top + G_2 q(t_k^m) G_2^\top) + \sum_{j=1}^{N-1} \beta(1-\beta)^j \mathbb{E}[(C(\tilde{A}^{-j}(\mathcal{X}(t_k^m) - \tilde{B} - \tilde{U}_w \tilde{\mathcal{W}}(t_k^{m-1})
\end{aligned}$$

$$- \tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-1}))\tilde{\mathcal{S}}_1(t_k^{m-1}) - (\hat{\mathcal{X}}(t_k^m|t_k) - \tilde{B})) +$$

$$\sum_{l=1}^{j-1} A^{-(j-l)}(-\tilde{B} - \tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-(l+1)}) - \tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-(l+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(l+1)})$$

$$+ \tilde{B})) + U_v\mathcal{V}(t_k^{m-j}) + G_2 \operatorname{diag}(\mathcal{X}(t_k^{m-j}))\mathcal{S}_2(t_k^{m-j}))(\star)^\top]$$

$$= \beta(C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top + G_2q(t_k^m)G_2^\top) + \sum_{j=1}^{N-1} \beta(1-\beta)^j(C(\tilde{A}^{-j}(\mathbb{E}[((\mathcal{X}(t_k^m) -$$

$$\hat{\mathcal{X}}(t_k^m|t_k)))(\star)^\top] + \tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-j})^\top$$

$$+ \sum_{l=1}^{j-1} \tilde{A}^{-(j-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-l)})^\top + \mathbb{E}[(\tilde{A}^{-j}(\mathcal{X}(t_k^m) - \hat{\mathcal{X}}(t_k^m|t_k)))$$

$$(\tilde{A}^{-j}(-\tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-1}) - \tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-1}))\tilde{\mathcal{S}}_1(t_k^{m-1})) + \sum_{l=1}^{j-1} \tilde{A}^{-(j-l)}(-\tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-(l+1)}) -$$

$$\tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-(l+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(l+1)})))^\top] + \mathbb{E}[(\tilde{A}^{-j}(-\tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-1}) - \tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-1}))\tilde{\mathcal{S}}_1(t_k^{m-1}))$$

$$+ \sum_{l=1}^{j-1} A^{-(j-l)}(-\tilde{B} - \tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-(l+1)}) - \tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-(l+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(l+1)})))$$

$$(\tilde{A}^{-j}(\mathcal{X}(t_k^m) - \hat{\mathcal{X}}(t_k^m|t_k)))^\top])C^\top + U_vU_v^\top + G_2q(t_k^{m-j})G_2^\top)$$

$$= \beta(C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top + G_2q(t_k^m)G_2^\top) + \sum_{j=1}^{N-1} \beta(1-\beta)^j(C(\tilde{A}^{-j}(\bar{P}(t_k^m|t_k) + \tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-j})^\top + \sum_{l=1}^{j-1} \tilde{A}^{-(j-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-l)})^\top - \tilde{A}^{-j}(\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-j})^\top - \sum_{l=1}^{j-1} \tilde{A}^{-(j-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-l)})^\top - \tilde{A}^{-j}(\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-j})^\top - \sum_{l=1}^{j-1} \tilde{A}^{-(j-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-l)})^\top)C^\top + U_vU_v^\top$$

$$+ G_2q(t_k^{m-j})G_2^\top)$$

$$= \beta(C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top + G_2q(t_k^m)G_2^\top) + \sum_{j=1}^{N-1} \beta(1-\beta)^j(C(\tilde{A}^{-j}\bar{P}(t_k^m|t_k)(\tilde{A}^{-j})^\top$$

$$- \tilde{A}^{-j}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-j})^\top - \sum_{l=1}^{j-1} \tilde{A}^{-(j-l)}(\tilde{U}_w\tilde{U}_w^\top$$

$$+ \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-l)})^\top)C^\top + U_vU_v^\top + G_2q(t_k^{m-j})G_2^\top)$$

$$= \beta(C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top + G_2q(t_k^m)G_2^\top) + \sum_{j=1}^{N-1}\beta(1-\beta)^j(C(\tilde{A}^{-j}\bar{P}(t_k^m|t_k)(\tilde{A}^{-j})^\top$$
$$- \sum_{l=0}^{j-1}\tilde{A}^{-(j-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-l)})^\top)C^\top + U_vU_v^\top + G_2q(t_k^{m-j})G_2^\top) \tag{A.48}$$

$$\mathbb{E}[T_2T_2^\top] = \mathbb{E}[(\prod_{j=0}^{N}p_{jk})(\mathcal{Y}(t_k^{m-N}) - \hat{\mathcal{Y}}(t_k^{m-N}|t_k))(\star)^\top] = \mathbb{E}[(\prod_{j=0}^{N}p_{jk})^2]\mathbb{E}[(\mathcal{Y}(t_k^{m-N}) - \hat{\mathcal{Y}}(t_k^{m-N}|t_k))(\star)^\top]$$
$$= (1-\beta)^N(C(\tilde{A}^{-N}\bar{P}(t_k^m|t_k)(\tilde{A}^{-N})^\top - \tilde{A}^{-N}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-N})^\top$$
$$- \sum_{l=1}^{N-1}\tilde{A}^{-(N-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(A^{-(N-l)})^\top)C^\top + U_vU_v^\top + G_2q(t_k^{m-N})G_2^\top) \tag{A.49}$$

$$\mathbb{E}[T_3T_3^\top] = \mathbb{E}[\sum_{j=0}^{N-1}\left((\prod_{i=0}^{j}p_{ik})(1-p_{(j+1)k}) - \beta(1-\beta)^j\right)\hat{\mathcal{Y}}(t_k^{m-j}|t_k)(\star)^\top]$$
$$= \mathbb{E}[\sum_{j=0}^{N-1}\left((\prod_{i=0}^{j}p_{ik})(1-p_{(j+1)k}) - \beta(1-\beta)^j\right)^2]\mathbb{E}[\hat{\mathcal{Y}}(t_k^{m-j}|t_k)\hat{\mathcal{Y}}(t_k^{m-j}|t_k)^\top]$$
$$= \sum_{j=0}^{N-1}(\beta(1-\beta)^j - \beta^2(1-\beta)^{2j})\psi_j(t_k^{m-j})\psi_j(t_k^{m-j})^\top \tag{A.50}$$

$$\mathbb{E}[T_4T_4^\top] = \mathbb{E}[\left((\prod_{j=0}^{N}p_{jk}) - (1-\beta)^N\right)\hat{\mathcal{Y}}(t_k^{m-N})|t_k)(\star)^\top]$$
$$= \mathbb{E}[((\prod_{j=0}^{N}p_{jk}) - (1-\beta)^N)^2]\mathbb{E}[\hat{\mathcal{Y}}(t_k^{m-N}|t_k)(\star)^\top]$$
$$= ((1-\beta)^N - (1-\beta)^{2N})\psi_N(t_k^N)\psi_N(t_k^N)^\top \tag{A.51}$$

$$\mathbb{E}(T_3T_4^\top) = -\beta(1-\beta)^N\sum_{j=0}^{N-1}(1-\beta)^j\psi_j(t_k^{m-j})\psi_N(t_k^N)^\top, \tag{A.52}$$

$$\mathbb{E}(T_4T_3^\top) = -\beta(1-\beta)^N\psi_N(t_k^N)\sum_{j=0}^{N-1}(1-\beta)^j\psi_j(t_k^{m-j})^\top, \tag{A.53}$$

where $\hat{\mathcal{Y}}(t_k^{m-j}|t_k) = \psi_j(t_k^{m-j})$.
Substituting (A.48) - (A.53) at (A.47) we have

$$\begin{aligned}
&\hat{\chi}(t_k) = \mathbb{E}[(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))(\star)^\top] \\
&= \beta(C\bar{P}(t_k^m|t_k)C^\top + U_vU_v^\top + G_2q(t_k^m)G_2^\top) + \sum_{j=1}^{N-1}\beta(1-\beta)^j(C(\tilde{A}^{-j}\bar{P}(t_k^m|t_k)(\tilde{A}^{-j})^\top \\
&- \sum_{l=0}^{j-1}\tilde{A}^{-(j-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(j-l)})^\top)C^\top + U_vU_v^\top + G_2q(t_k^{m-j})G_2^\top) \\
&+ (1-\beta)^N(C(\tilde{A}^{-N}\bar{P}(t_k^m|t_k)(\tilde{A}^{-N})^\top - \tilde{A}^{-N}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-1})\tilde{G}_1^\top)(\tilde{A}^{-N})^\top \\
&- \sum_{l=1}^{N-1}\tilde{A}^{-(N-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(l+1)})\tilde{G}_1^\top)(A^{-(N-l)})^\top)C^\top + U_vU_v^\top + G_2q(t_k^{m-N})G_2^\top) + \\
&\sum_{j=0}^{N-1}(\beta(1-\beta)^j - \beta^2(1-\beta)^{2j})\psi_j(t_k^{m-j})\psi_j(t_k^{m-j})^\top + ((1-\beta)^N - (1-\beta)^{2N})\psi_N(t_k^N)\psi_N(t_k^N)^\top \\
&- \beta(1-\beta)^N\sum_{j=0}^{N-1}(1-\beta)^j\psi_j(t_k^{m-j})\psi_N(t_k^N)^\top - \beta(1-\beta)^N\psi_N(t_k^N)\sum_{j=0}^{N-1}(1-\beta)^j\psi_j(t_k^{m-j})^\top.
\end{aligned} \tag{A.54}$$

Next, we need the following covariance terms in evaluating $\bar{P}(t_k^{i+1}|t_{k+1})$:

$$\begin{aligned}
&\mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))^\top] \\
&= \mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))(\sum_{j=0}^{N-1}(\prod_{i=0}^{j}p_{ik})(1-p_{(j+1)k}) \\
&(\mathcal{Y}(t_k^{m-j}) - \hat{\mathcal{Y}}(t_k^{m-j}|t_k)) + (\prod_{j=0}^{N}p_{jk})(\mathcal{Y}(t_k^{m-N}) - \hat{\mathcal{Y}}(t_k^{m-N}|t_k)) + \\
&\sum_{j=0}^{N-1}\left((\prod_{i=0}^{j}p_{ik})(1-p_{(j+1)k}) - \beta(1-\beta)^j\right)\hat{\mathcal{Y}}(t_k^{m-j}|t_k) + \left((\prod_{j=0}^{N}p_{jk}) - (1-\beta)^N\right)\hat{\mathcal{Y}}(t_k^{m-N}|t_k))^\top]
\end{aligned}$$

$$= \mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))$$

$$(\sum_{j=0}^{N-1}(\prod_{i=0}^{j} p_{ik})(1-p_{(j+1)k})(\mathcal{Y}(t_k^{m-j}) - \hat{\mathcal{Y}}(t_k^{m-j}|t_k)))^\top]$$

$$+ \mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))((\prod_{j=0}^{N} p_{jk})(\mathcal{Y}(t_k^{m-N}) - \hat{\mathcal{Y}}(t_k^{m-N}|t_k)))^\top]$$

$$+ \mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))$$

$$\underbrace{(\sum_{j=0}^{N-1}\left((\prod_{i=0}^{j} p_{ik})(1-p_{(j+1)k}) - \beta(1-\beta)^j\right)}_{0}\hat{\mathcal{Y}}(t_k^{m-j}|t_k))^\top]$$

$$+ \mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))\underbrace{(\left((\prod_{j=0}^{N} p_{jk}) - (1-\beta)^N\right)}_{0}\hat{\mathcal{Y}}(t_k^{m-N}|t_k))^\top]$$

$$= \sum_{j=0}^{N-1}\beta(1-\beta)^j\mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))$$

$$(C(\mathcal{X}(t_k^{m-j}) - \hat{\mathcal{X}}(t_k^{m-j}|t_k)) + U_v\mathcal{V}(t_k^{m-j}) + G_2\mathcal{X}(t_k^{m-j})\mathcal{S}_2(t_k^{m-j}))))^\top]+$$

$$(1-\beta)^N\mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))$$

$$(C(\mathcal{X}(t_k^{m-N}) - \hat{\mathcal{X}}(t_k^{m-N}|t_k)) + U_v\mathcal{V}(t_k^{m-N}) + G_2\mathcal{X}(t_k^{m-N})\mathcal{S}_2(t_k^{m-N}))))^\top]$$

$$= \sum_{j=0}^{N-1}\beta(1-\beta)^j\mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))(C(\mathcal{X}(t_k^{m-j}) - \hat{\mathcal{X}}(t_k^{m-j}|t_k))))^\top]$$

$$+ (1-\beta)^N\mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m\tilde{\mathcal{W}}(t_k^i) + \tilde{G}_1\mathcal{X}(t_k^i)\tilde{\mathcal{S}}_1(t_k^i))(C((\mathcal{X}(t_k^{m-N}) - \hat{\mathcal{X}}(t_k^{m-N}|t_k))))^\top]$$

$$= \beta(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)C^\top - \sum_{r=0}^{m-i-2}\tilde{A}^{-(m-i-r)+1}((\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^r)^\top C^\top)$$

$$+ \sum_{j=1}^{N-1}\beta(1-\beta)^j\mathbb{E}[\tilde{A}(\tilde{A}^{i-m}(\mathcal{X}(t_k^m) - \tilde{B} - (\hat{\mathcal{X}}(t_k^m|t_k) - \tilde{B}))+$$

$$\sum_{r=0}^{m-i-2}\tilde{A}^{-(m-i-r)}(-\tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-(r+1)}) - \tilde{G}_1\operatorname{diag}(\mathcal{X}_j(t_k^{m-(r+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(r+1)})))$$

$$(C(\tilde{A}^{-j}(\mathcal{X}(t_k^m) - \tilde{B} - (\hat{\mathcal{X}}(t_k^m|t_k) - \tilde{B}))+$$

$$\sum_{l=0}^{j-1}A^{-(j-l)}(-\tilde{B} - \tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-(l+1)}) - \tilde{G}_1\operatorname{diag}(\mathcal{X}(t_k^{m-(l+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(l+1)}) + \tilde{B})))^\top]+$$

$$(1-\beta)^N \mathbb{E}[(\tilde{A}(\tilde{A}^{i-m}(\mathcal{X}(t_k^m)-\tilde{B})-(\hat{\mathcal{X}}(t_k^m|t_k)-\tilde{B}))+$$

$$\sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)}(-\tilde{U}_m \tilde{\mathcal{W}}(t_k^{m-(r+1)})-\tilde{G}_1 \operatorname{diag}(\mathcal{X}_j(t_k^{m-(r+1)}))$$

$$\tilde{\mathcal{S}}_1(t_k^{m-(r+1)}))))(C(\tilde{A}^{-N}(\mathcal{X}(t_k^m)-\tilde{B}-(\hat{\mathcal{X}}(t_k^m|t_k)-\tilde{B}))+$$

$$\sum_{l=0}^{N-1} \tilde{A}^{-(N-l)}(-\tilde{U}_w \tilde{\mathcal{W}}(t_k^{m-(l+1)}))-\tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-(l+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(l+1)})))^\top]$$

$$=\beta(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)C^\top - \sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}((\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^r)^\top C^\top)$$

$$+\sum_{j=1}^{N-1} \beta(1-\beta)^j \tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k)\tilde{A}^{-j}C^\top + \sum_{s=0}^{min\{j-1,m-i-2\}} A^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top$$

$$+\tilde{G}_1 q(t_k^{m-(s+1)})\tilde{G}_1^\top)(A^{-(j-s)})^\top C^\top + \mathbb{E}[(\tilde{A}^{i-m+1}\mathcal{X}(t_k^m))(C\sum_{l=0}^{j-1} A^{-(j-l)}(-\tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-(l+1)})$$

$$-\tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-(l+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(l+1)}))))^\top] + \mathbb{E}[(\sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(-\tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-(r+1)})$$

$$-\tilde{G}_1 \operatorname{diag}(\mathcal{X}_j(t_k^{m-(r+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(r+1)})))(C\tilde{A}^{-j}(\mathcal{X}(t_k^m))^\top] + (1-\beta)^N(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)\tilde{(}A^{-N})^\top C^\top +$$

$$\sum_{s=0}^{min\{N-1,m-i-2\}} A^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1)})\tilde{G}_1^\top)(A^{-(N-s)})^\top C^\top +$$

$$\mathbb{E}[(A^{i-m+1}\mathcal{X}(t_k^m))(C(\sum_{l=0}^{N-1}\tilde{A}^{-(N-l)}(-\tilde{U}_w\tilde{\mathcal{W}}(t_k^{m-(l+1)}))-\tilde{G}_1 \operatorname{diag}(\mathcal{X}(t_k^{m-(l+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(l+1)}))))^\top]+$$

$$\mathbb{E}[(\sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(-\tilde{U}_m\tilde{\mathcal{W}}(t_k^{m-(r+1)})-\tilde{G}_1 \operatorname{diag}(\mathcal{X}_j(t_k^{m-(r+1)}))\tilde{\mathcal{S}}_1(t_k^{m-(r+1)})))(C\tilde{A}^{-N}(\mathcal{X}(t_k^m)))^\top]$$

$$=\beta(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)C^\top - \sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}((\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^r)^\top C^\top)$$

$$+\sum_{j=1}^{N-1} \beta(1-\beta)^j \tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k)(\tilde{A}^{-j})^\top C^\top + \sum_{s=0}^{min\{j-1,m-i-2\}} A^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top$$

$$+\tilde{G}_1 q(t_k^{m-(s+1)})\tilde{G}_1^\top)(A^{-(j-s)})^\top C^\top - \sum_{l=0}^{j-1} \tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1)})(\tilde{A}^{-(j-l)})^\top C^\top$$

$$- \sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}((\tilde{U}_w \tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1))}) \tilde{G}_1^\top)(\tilde{A}^{-(j-r)})^\top C^\top) +$$

$$(1-\beta)^N (\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k)(\tilde{A}^{-N})^\top C^\top + \sum_{s=0}^{min\{N-1,m-i-2\}} \tilde{A}^{-(m-i-s)+1}(\tilde{U}_w \tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(s+1))}) \tilde{G}_1^\top)(A^{-(N-s)})^\top C^\top - \sum_{l=0}^{N-1} \tilde{A}^{-(m-i-l)+1}(\tilde{U}_w \tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1)}) \tilde{G}_1^\top)(\tilde{A}^{-(N-l)})^\top C^\top$$

$$- \sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_w \tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1)}) \tilde{G}_1^\top)(\tilde{A}^{-(N-r)})^\top C^\top \tag{A.55}$$

and

$$\mathbb{E}[(\mathcal{Z}(t_{k+1}) - \hat{\mathcal{Z}}(t_{k+1}|t_k))(\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k)) + \tilde{U}_m \tilde{\mathcal{W}}(t_k^{i+1}) + \tilde{G}_1 \mathcal{X}(t_k^i) \tilde{\mathcal{S}}_1(t_k^{i+1}))^\top]$$

$$= \beta(C\bar{P}(t_k^m|t_k)(\tilde{A}^{i-m+1})^\top - \sum_{r=0}^{m-i-2} C\tilde{A}^r(\tilde{U}_w \tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1))} \tilde{G}_1^\top)(\tilde{A}^{-(m-i-r)+1})^\top) +$$

$$\sum_{j=1}^{N-1} \beta(1-\beta)^j (C\tilde{A}^{-j}(\bar{P}(t_k^m|t_k)(\tilde{A}^{i-m+1})^\top + \sum_{s=0}^{min\{j-1,m-i-2\}} C\tilde{A}^{-(j-s)}(\tilde{U}_w \tilde{U}_w^\top$$

$$+ \tilde{G}_1 q(t_k^{m-(s+1))}) \tilde{G}_1^\top)(\tilde{A}^{-(m-i-s)+1})^\top - \sum_{l=0}^{j-1} C\tilde{A}^{-(j-l)}(\tilde{U}_w \tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1))})(\tilde{A}^{-(m-i-l)+1})^\top$$

$$- \sum_{r=0}^{m-i-2} C\tilde{A}^{-(j-r)}((\tilde{U}_w \tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1))}) \tilde{G}_1^\top)(\tilde{A}^{-(m-i-r)+1})^\top)$$

$$+ (1-\beta)^N (C\tilde{A}^{-N}(\bar{P}(t_k^m|t_k)(\tilde{A}^{i-m+1})^\top + \sum_{s=0}^{min\{N-1,m-i-2\}} CA^{-(N-s)}(\tilde{U}_w \tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(s+1))}) \tilde{G}_1^\top)(\tilde{A}^{-(m-i-s)+1})^\top - \sum_{l=0}^{N-1} C\tilde{A}^{-(N-l)}(\tilde{U}_w \tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(l+1)}) \tilde{G}_1^\top)(\tilde{A}^{-(m-i-l)+1})^\top - \sum_{r=0}^{m-i-2} C\tilde{A}^{-(N-r)}(\tilde{U}_w \tilde{U}_w^\top$$

$$+ \tilde{G}_1 q(t_k^{m-(r+1)}) \tilde{G}_1^\top)(\tilde{A}^{-(m-i-r)+1})^\top) \tag{A.56}$$

Substituting (A.54)- (A.56) at (A.42) we have

$$\bar{P}(t_k^{i+1}|t_{k+1}) = \bar{P}(t_k^{i+1}|t_k) + \bar{K}(t_{k+1}^i)\hat{\chi}(t_k)\bar{K}(t_{k+1}^i)^\top - (\beta(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)C^\top -$$

$$\sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}((\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1))})\tilde{G}_1^\top)(\tilde{A}^r)^\top C^\top) + \sum_{j=1}^{N-1} \beta(1-\beta)^j \tilde{A}^{i-m+1} \times$$

$$(\bar{P}(t_k^m|t_k)(\tilde{A}^{-j})^\top C^\top + \sum_{s=0}^{min\{j-1,m-i-2\}} A^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1))})\tilde{G}_1^\top)(A^{-(j-s)})^\top C^\top -$$

$$\sum_{l=0}^{j-1} \tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1))})(\tilde{A}^{-(j-l)})^\top C^\top - \sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}((\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(r+1))})\tilde{G}_1^\top)(\tilde{A}^{-(j-r)})^\top C^\top) + (1-\beta)^N(\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k)(\tilde{A}^{-N})^\top C^\top +$$

$$\sum_{s=0}^{min\{N-1,m-i-2\}} \tilde{A}^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1))})\tilde{G}_1^\top)(A^{-(N-s)})^\top C^\top -$$

$$\sum_{l=0}^{N-1} \tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-l)})^\top C^\top -$$

$$\sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-r)})^\top C^\top))\bar{K}(t_{k+1}^i)^\top -$$

$$\bar{K}(t_{k+1}^i)(\beta(C\bar{P}(t_k^m|t_k)(\tilde{A}^{i-m+1})^\top - \sum_{r=0}^{m-i-2} C\tilde{A}^r(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1))}\tilde{G}_1^\top)(\tilde{A}^{-(m-i-r)+1})^\top)$$

$$+ \sum_{j=1}^{N-1} \beta(1-\beta)^j(C\tilde{A}^{-j}(\bar{P}(t_k^m|t_k)(\tilde{A}^{i-m+1})^\top + \sum_{s=0}^{min\{j-1,m-i-2\}} C\tilde{A}^{-(j-s)}(\tilde{U}_w\tilde{U}_w^\top$$

$$+ \tilde{G}_1 q(t_k^{m-(s+1))})\tilde{G}_1^\top)(\tilde{A}^{-(m-i-s)+1})^\top - \sum_{l=0}^{j-1} C\tilde{A}^{-(j-l)}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1))})(\tilde{A}^{-(m-i-l)+1})^\top$$

$$- \sum_{r=0}^{m-i-2} C\tilde{A}^{-(j-r)}((\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1))})\tilde{G}_1^\top)(\tilde{A}^{-(m-i-r)+1})^\top)+$$

$$(1-\beta)^N(C\tilde{A}^{-N}(\bar{P}(t_k^m|t_k)(\tilde{A}^{i-m+1})^\top + \sum_{s=0}^{min\{N-1,m-i-2\}} CA^{-(N-s)}(\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(s+1))})\tilde{G}_1^\top)(\tilde{A}^{-(m-i-s)+1})^\top - \sum_{l=0}^{N-1} C\tilde{A}^{-(N-l)}(\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(m-i-l)+1})^\top - \sum_{r=0}^{m-i-2} C\tilde{A}^{-(N-r)}(\tilde{U}_w\tilde{U}_w^\top$$

$$+ \tilde{G}_1 q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^{-(m-i-r)+1})^\top) \tag{A.57}$$

where

$$\bar{P}(t_k^{i+1}|t_k) = \tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top + \tilde{U}_m\tilde{U}_m^\top + \tilde{G}_1 q(t_k^i)\tilde{G}_1^\top,$$

$\bar{P}(t_k^m|t_k)$ is as defined in equation (4.26).

To find the value of $\bar{K}(k+1)$ that minimizes the trace of the estimation error covariance matrix $\bar{P}(t_k^{i+1}|t_{k+1})$ we differentiate the trace of the above expression with respect to the filter gain matrix $\bar{K}(k+1)$ and set the derivative to zero.

$$\frac{\partial tr\bar{P}(t_k^{i+1}|t_{k+1})}{\partial \bar{K}(t_{k+1}^i)} = -2(\beta(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)C^\top - \sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}((\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(r+1))})\tilde{G}_1^\top)(\tilde{A}^r)^\top C^\top) + \sum_{j=1}^{N-1} \beta(1-\beta)^j(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)(\tilde{A}^{-j})^\top C^\top +$$

$$\sum_{s=0}^{min\{j-1,m-i-2\}} A^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1))}\tilde{G}_1^\top)(A^{-(j-s)})^\top C^\top -$$

$$\sum_{l=0}^{j-1} \tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1))})(\tilde{A}^{-(j-l)})^\top C^\top - \sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(r+1))}\tilde{G}_1^\top)(\tilde{A}^{-(j-r)})^\top C^\top) + (1-\beta)^N(\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k)(\tilde{A}^{-N})^\top C^\top +$$

$$\sum_{s=0}^{min\{N-1,m-i-2\}} \tilde{A}^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1))})\tilde{G}_1^\top)(A^{-(N-s)})^\top C^\top -$$

$$\sum_{l=0}^{N-1} \tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-l)})^\top C^\top -$$

$$\sum_{r=0}^{m-i-2} \tilde{A}^{-(m-i-r)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-r)})^\top C^\top) + 2\bar{K}(t_{k+1}^i)\hat{\chi}(t_k). \tag{A.58}$$

Then

$$\bar{K}(t_{k+1}^{i}) = (\beta(\tilde{A}^{i-m+1}\bar{P}(t_k^m|t_k)C^\top - \sum_{r=0}^{m-i-2}\tilde{A}^{-(m-i-r)+1}((\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(r+1))})\tilde{G}_1^\top)(\tilde{A}^r)^\top C^\top) + \sum_{j=1}^{N-1}\beta(1-\beta)^j(\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k)(\tilde{A}^{-j})^\top C^\top +$$

$$\sum_{s=0}^{min\{j-1,m-i-2\}} A^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1))})\tilde{G}_1^\top)(A^{-(j-s)})^\top C^\top -$$

$$\sum_{l=0}^{j-1}\tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1))})(\tilde{A}^{-(j-l)})^\top C^\top - \sum_{r=0}^{m-i-2}\tilde{A}^{-(m-i-r)+1}((\tilde{U}_w\tilde{U}_w^\top +$$

$$\tilde{G}_1 q(t_k^{m-(r+1))})\tilde{G}_1^\top)(\tilde{A}^{-(j-r)})^\top C^\top) + (1-\beta)^N(\tilde{A}^{i-m+1}(\bar{P}(t_k^m|t_k)(\tilde{A}^{-N})^\top C^\top +$$

$$\sum_{s=0}^{min\{N-1,m-i-2\}}\tilde{A}^{-(m-i-s)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(s+1))})\tilde{G}_1^\top)(A^{-(N-s)})^\top C^\top -$$

$$\sum_{l=0}^{N-1}\tilde{A}^{-(m-i-l)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(l+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-l)})^\top C^\top -$$

$$\sum_{r=0}^{m-i-2}\tilde{A}^{-(m-i-r)+1}(\tilde{U}_w\tilde{U}_w^\top + \tilde{G}_1 q(t_k^{m-(r+1)})\tilde{G}_1^\top)(\tilde{A}^{-(N-r)})^\top C^\top)\hat{\chi}(t_k)^{-1} \quad \text{(A.59)}$$

which is the required result. ∎

Proof of Theorem 6

The proof follows along the same lines as the proof of earlier theorems. The filtering estimates of the state covariance is obtained by combining the equations (5.1)-(5.5) as follows

$$\bar{P}(k+1|k+1) = \mathbb{E}[f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k))(\star)^\top] + U_w U_w^\top +$$
$$\bar{K}(k+1)(\mathbb{E}[(1-p_{k+1})^2]\mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^\top]$$
$$+ \mathbb{E}[(p_{k+1})^2]\mathbb{E}[(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k))(\star)^\top] + \mathbb{E}[((1-p_{k+1}) - \beta)^2]$$
$$(\hat{\mathcal{Y}}(k+1|k))(\star)^\top + \mathbb{E}[(p_{k+1} - (1-\beta))^2](\hat{\mathcal{Y}}(k|k))(\star)^\top$$

$$
\begin{aligned}
&+ \underbrace{\mathbb{E}[(1-p_{k+1})(\mathcal{Y}(k+1)-\hat{\mathcal{Y}}(k+1|k))(p_{k+1}(\mathcal{Y}(k)-\hat{\mathcal{Y}}(k|k))^\top]}_{\text{equals zero}} \\
&+ \underbrace{\mathbb{E}[p_{k+1}(\mathcal{Y}(k)-\hat{\mathcal{Y}}(k|k))(1-p_{k+1})(\mathcal{Y}(k+1)-\hat{\mathcal{Y}}(k+1|k))^\top]}_{\text{equals zero}} \\
&+ \mathbb{E}[((1-p_{k+1})-\beta)(p_{k+1}-(1-\beta))(\hat{\mathcal{Y}}(k+1|k)\hat{\mathcal{Y}}(k|k)^\top + \\
&\hat{\mathcal{Y}}(k|k)\hat{\mathcal{Y}}(k+1|k)^\top)])\bar{K}(k+1)^\top - \bar{K}(k+1)(\mathbb{E}[(f(\mathcal{X}(k)- \\
&f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k)((1-p_{k+1})(\mathcal{Y}(k+1)-\hat{\mathcal{Y}}(k+1|k))+ \\
&p_{k+1}(\mathcal{Y}(k)-\hat{\mathcal{Y}}(k|k)) + ((1-p_{k+1})-\beta)\hat{\mathcal{Y}}(k+1|k)+ \\
&(p_{k+1}-(1-\beta))\hat{\mathcal{Y}}(k|k))^\top] + \mathbb{E}[((1-p_{k+1})(\mathcal{Y}(k+1)- \\
&\hat{\mathcal{Y}}(k+1|k)) + p_{k+1}(\mathcal{Y}(k)-\hat{\mathcal{Y}}(k|k)) + ((1-p_{k+1})-\beta) \\
&\hat{\mathcal{Y}}(k+1|k) + (p_{k+1}-(1-\beta))\hat{\mathcal{Y}}(k|k)) \\
&(f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k). \qquad \text{(A.60)}
\end{aligned}
$$

We can derive the following easily:

$$
\begin{aligned}
&\mathbb{E}[f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k))(\star)^\top] \\
&= \mathbb{E}[f(\hat{\mathcal{X}}(k|k)) + A(k)\tilde{\mathcal{X}}(k|k) + Q_1(k)\tilde{\mathcal{X}}(k|k)R_1(k) - f(\hat{\mathcal{X}}(k|k))(\star)^\top] \\
&= A(k)\bar{P}(k|k)A(k)^\top + Q_1(k)\bar{P}(k|k)Q_1(k)^\top
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}[h(\mathcal{X}(k+1)) - h(\hat{\mathcal{X}}(k+1|k))(\star)^\top] \\
&= \mathbb{E}[(h(\hat{\mathcal{X}}(k+1|k)) + C(k+1)\tilde{\mathcal{X}}(k+1|k) + Q_2\tilde{\mathcal{X}}(k+1|k)R_2(k) - h(\hat{\mathcal{X}}(k+1|k)))(\star)^\top] \\
&= (C(k+1)\bar{P}(k+1|k)(C(k+1)^\top + Q_2(k+1)\bar{P}(k+1|k)Q_2(k+1)^\top, \qquad \text{(A.61)}
\end{aligned}
$$

Next, we need (A.9) and the following covariance terms in evaluating $\bar{P}(k+1|k+1)$:

$$
\begin{aligned}
&\mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^\top] \\
&= C(k+1)\bar{P}(k+1|k)C(k+1)^\top + Q_2(k+1)\bar{P}(k+1|k)Q_2(k+1)^\top + U_vU_v^\top, \\
&\mathbb{E}[(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k))(\star)^\top] \\
&= C(k)\bar{P}(k|k)C(k)^\top + Q_2(k)\bar{P}(k|k)Q_2(k)^\top + U_vU_v^\top \\
&\text{where} \quad \bar{P}(k+1|k) = A(k)\bar{P}(k|k)A(k)^\top + Q_1(k)\bar{P}(k|k)Q_1(k) + U_wU_w^\top. \qquad \text{(A.62)}
\end{aligned}
$$

and $\hat{\mathcal{Y}}(k+i|k) = h(\hat{\mathcal{X}}(k+i|k)), i = 0, 1$. For further notational brevity, denote

$$\psi_i(k+1) = \hat{\mathcal{Y}}(k+i|k), i = 0,1$$
$$\text{and } \widetilde{\psi}_i(k+1) := \psi_i(k+1)\psi_i(k+1)^\top. \tag{A.63}$$

We also need to evaluate some cross covariance terms, whose expressions are derived next:

$$\begin{aligned}
&\mathbb{E}[(f(\mathcal{X}(k) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k))((1-p_{k+1})(\mathcal{Y}(k+1)- \\
&\hat{\mathcal{Y}}(k+1|k)) + p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k))+ \\
&((1-p_{k+1}) - \beta)\hat{\mathcal{Y}}(k+1|k) + (p_{k+1} - (1-\beta))\hat{\mathcal{Y}}(k|k))^\top] \\
&= \mathbb{E}[(f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k))((1-p_{k+1})(\mathcal{Y}(k+1) \\
&- \hat{\mathcal{Y}}(k+1|k)))^\top] + \mathbb{E}[(f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k)) \\
&(p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)))^\top] + \mathbb{E}[(f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k)) \\
&((1-p_{k+1}) - \beta)\hat{\mathcal{Y}}(k+1|k)^\top] + \mathbb{E}[(f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k))+ \\
&U_w\mathcal{W}(k))(p_{k+1} - (1-\beta))\hat{\mathcal{Y}}(k|k))^\top]
\end{aligned} \tag{A.64}$$

$$\begin{aligned}
&\mathbb{E}[(f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k))((1-p_{k+1}) \\
&(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)))^\top] \\
&= \mathbb{E}[1-p_{k+1}]\mathbb{E}[(f(\hat{\mathcal{X}}(k|k)) + A(k)\tilde{\mathcal{X}}(k|k) + Q_1(k)\tilde{\mathcal{X}}(k|k) \\
&\times R_1(k) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k))(h(\hat{\mathcal{X}}(k+1|k))+ \\
&C(k+1)\tilde{\mathcal{X}}(k+1|k) + Q_2(k+1)\tilde{\mathcal{X}}(k+1|k)R_2(k+1) \\
&+ U_v\mathcal{V}(k+1) - h(\hat{\mathcal{X}}(k+1|k)))^\top] \\
&= \beta\mathbb{E}[(A(k)\tilde{\mathcal{X}}(k|k)) + Q_1(k)\tilde{\mathcal{X}}(k|k)R_1(k) + U_w\mathcal{W}(k))\times \\
&(C(k+1)(A(k)\tilde{\mathcal{X}}(k|k) + Q_1(k)\tilde{\mathcal{X}}(k|k)R_1(k) + U_w\mathcal{W}(k))+ \\
&Q_2(k+1)\tilde{\mathcal{X}}(k+1|k)R_2(k+1) + U_v\mathcal{V}(k+1))^\top] \\
&= \beta(A(k)\bar{P}(k|k)A(k)^\top + Q_1(k)\bar{P}(k|k)Q_1(k)^\top + U_wU_w^\top)C(k+1)^\top \\
&= \beta\bar{P}(k+1|k)C(k+1)^\top
\end{aligned} \tag{A.65}$$

and

$$
\begin{aligned}
&\mathbb{E}[(f(\mathcal{X}(k)) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k))(p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)))^\top] \\
&= (1-\beta)\mathbb{E}[(f(\hat{\mathcal{X}}(k|k)) + A(k)\tilde{\mathcal{X}}(k|k) + Q_1(k)\tilde{\mathcal{X}}(k|k)R_1(k) + \\
&U_w\mathcal{W}(k) - f(\hat{\mathcal{X}}(k|k)))(h(\hat{\mathcal{X}}(k|k)) + C(k)\tilde{\mathcal{X}}(k|k) + Q_2(k)\tilde{\mathcal{X}}(k|k)R_2(k) \\
&+ U_v\mathcal{V}(k) - h(\hat{\mathcal{X}}(k|k)))^\top)] \\
&= (1-\beta)A(k)\bar{P}(k|k)C(k)^\top
\end{aligned} \tag{A.66}
$$

Substituting (A.65) and (A.66) in (A.64)

$$
\begin{aligned}
&\mathbb{E}[(f(\mathcal{X}(k) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k)) \\
&((1-p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)) + p_{k+1}(\mathcal{Y}(k) - \hat{\mathcal{Y}}(k|k)) \\
&+ ((1-p_{k+1}) - \beta)\hat{\mathcal{Y}}(k+1|k) + (p_{k+1} - (1-\beta))\hat{\mathcal{Y}}(k|k))^\top] \\
&= \beta\bar{P}(k+1|k)C(k+1)^\top + (1-\beta)A(k)\bar{P}(k|k)C(k)^\top.
\end{aligned} \tag{A.67}
$$

Similarly,

$$
\begin{aligned}
&\mathbb{E}[((1-p_{k+1})(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)) + p_{k+1}(\mathcal{Y}(k) - \\
&\hat{\mathcal{Y}}(k|k)) + ((1-p_{k+1}) - \beta)\hat{\mathcal{Y}}(k+1|k) + (p_{k+1} - (1-\beta)) \\
&\hat{\mathcal{Y}}(k|k))(f(\mathcal{X}(k) - f(\hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k))^\top] \\
&= \beta C(k+1)\bar{P}(k+1|k) + (1-\beta)C(k)\bar{P}(k|k)A(k)^\top
\end{aligned} \tag{A.68}
$$

Substituting (A.62)-(A.68) in (A.60), we have

$$
\begin{aligned}
&\bar{P}(k+1|k+1) = A(k)\bar{P}(k|k)A(k)^\top + Q_1(k)\bar{P}(k|k)Q_1(k)^\top + \\
&U_wU_w^\top + \bar{K}(k+1)(\beta(C(k+1)\bar{P}(k+1|k)C(k+1)^\top + \\
&Q_2(k+1)\bar{P}(k+1|k)Q_2(k+1)^\top + U_vU_v^\top) + \\
&(1-\beta)(C(k)\bar{P}(k|k)C(k)^\top + Q_2(k)\bar{P}(k|k)Q_2(k)^\top + U_vU_v^\top) + \\
&\beta(1-\beta)(\widetilde{\psi}_0(k+1) + \widetilde{\psi}_1(k+1)) - \beta(1-\beta)(\psi_0(k+1)\psi_1(k+1)^\top \\
&+ \psi_1(k+1)\psi_0(k+1)^\top))\bar{K}(k+1)^\top - (\beta\bar{P}(k+1|k)C(k+1)^\top \\
&+ (1-\beta)A(k)\bar{P}(k|k)C(k)^\top)\bar{K}(k+1)^\top - \\
&\bar{K}(k+1)(\beta C(k+1)\bar{P}(k+1|k) + (1-\beta)C(k)\bar{P}(k|k)A(k)^\top)
\end{aligned}
$$

To find the value of $\bar{K}(k+1)$ that minimizes the trace of the covariance $\bar{P}(k+1|k+1)$ we differentiate the trace of the above expression with respect to the filter gain matrix $\bar{K}(k+1)$and set the derivative to zero.

$$\begin{aligned}
&\bar{K}(k+1) = (\beta \bar{P}(k+1|k)C(k+1)^\top + (1-\beta)A(k)\bar{P}(k|k)C(k)^\top) \\
&\times [\beta(C(k+1)\bar{P}(k+1|k)C(k+1)^\top + Q_2(k+1)\bar{P}(k+1|k)Q_2(k+1)^\top + U_v U_v^\top)+ \\
&(1-\beta)(C(k)\bar{P}(k|k)C(k)^\top + Q_2(k)\bar{P}(k|k)Q_2(k)^\top + U_v U_v^\top) + \beta(1-\beta)(\widetilde{\psi}_0(k+1) \\
&+ \widetilde{\psi}_1(k+1)) - \beta(1-\beta)(\psi_0(k+1)\psi_1(k+1)^\top + \psi_1(k+1)\psi_0(k+1)^\top)]^{-1}
\end{aligned} \tag{A.69}$$

which is the required expression. ∎

Bibliography

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of ASME (Journal of Basic Engineering)*, vol. 82, no. 1, pp. 35–45, 1960.

[2] G. Kallianpur, *Stochastic filtering theory*, vol. 13. Springer Science & Business Media, 2013.

[3] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Journal of basic engineering*, vol. 83, no. 1, pp. 95–108, 1961.

[4] S.-G. Kim, J. L. Crassidis, Y. Cheng, A. M. Fosbury, and J. L. Junkins, "Kalman filtering for relative spacecraft attitude and position estimation," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 1, pp. 133–143, 2007.

[5] M. Grewal and A. Andrews, *Kalman filtering: theory and practice using MATLAB*. John Wiley and Sons, 2008.

[6] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*, vol. 1. Prentice Hall Upper Saddle River, NJ, 2000.

[7] M. S. Grewal, *Kalman filtering*. Springer, 2011.

[8] B. Anderson and J. B. Moore, "Optimal filtering," *Prentice-Hall Information and System Sciences Series, Englewood Cliffs: Prentice-Hall, 1979*, 1979.

[9] P. Date, L. Jalen, and R. Mamon, "A new algorithm for latent state estimation in non-linear time series models," *Applied Mathematics and Computation*, vol. 203, no. 1, pp. 224–232, 2008.

[10] S. S. Haykin, *Kalman filtering and neural networks.* Wiley, 2001.

[11] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[12] P. L. Houtekamer and H. L. Mitchell, "Data assimilation using an ensemble Kalman filter technique," *Monthly Weather Review*, vol. 126, no. 3, pp. 796–811, 1998.

[13] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-time nonlinear filtering algorithms using gauss-hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[14] R. Kandepu, B. Foss, and L. Imsland, "Applying the unscented Kalman filter for nonlinear state estimation," *Journal of process control*, vol. 18, no. 7, pp. 753–768, 2008.

[15] J. L. Crassidis and J. L. Junkins, *Optimal estimation of dynamic systems.* Goong Chen and Thomas J Bridges, Eds. New York, USA: Chapman & Hall/CRC, 2004.

[16] L. Ljung, "Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems," *IEEE Transactions on Automatic Control*, vol. 24, no. 1, pp. 36–50, 1979.

[17] D. T. Pham, J. Verron, and M. C. Roubaud, "A singular evolutive extended Kalman filter for data assimilation in oceanography," *Journal of Marine systems*, vol. 16, no. 3, pp. 323–340, 1998.

[18] J. J. LaViola, "A comparison of unscented and extended Kalman filtering for estimating quaternion motion," in *Proceedings of the 2003 American Control Conference,*, vol. 3, pp. 2435–2440, IEEE, 2003.

[19] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proceedings of the 1995 American Control Conference,*, vol. 3, pp. 1628–1632, IEEE, 1995.

[20] E. A. Wan and R. van der Merwe, "The unscented Kalman filter for nonlinear estimating," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pp. 153–158, 2000.

[21] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Int. symp. aerospace/defense sensing, simul. and controls*, vol. 3, pp. 182–193, Orlando, FL, 1997.

[22] P. Date and K. Ponomareva, "Linear and non-linear filtering in mathematical finance: a review," *IMA Journal of Management Mathematics*, vol. 22, no. 3, pp. 195–211, 2010.

[23] A. UmaMageswari, J. J. Ignatious, and R. Vinodha, "A comparitive study Of Kalman filter, extended Kalman filter And unscented Kalman filter For harmonic analysis of the non-stationary signals ," *International Journal of Scientific & Engineering Research*, vol. 3, no. 7, pp. 1–9, 2012.

[24] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on automatic control*, vol. 45, no. 3, pp. 477–482, 2000.

[25] R. Van Der Merwe, E. A. Wan, S. Julier, *et al.*, "Sigma-point Kalman filters for nonlinear estimation and sensor-fusion: Applications to integrated navigation," in *Proceedings of the AIAA Guidance, Navigation & Control Conference*, pp. 16–19, 2004.

[26] K. Ponomareva, P. Date, and Z. Wang, "A new unscented Kalman filter with higher order moment-matching," *Proceedings of Mathematical Theory of Networks and Systems (MTNS 2010), Budapest*, vol. 11, pp. 1609–1613, 2010.

[27] T. S. Schei, "A finite-difference method for linearization in nonlinear estimation algorithms," *Automatica*, vol. 33, no. 11, pp. 2053–2058, 1997.

[28] A. Hermoso-Carazo and J. Linares-Pérez, "Different approaches for state filtering in nonlinear systems with uncertain observations," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 708–724, 2007.

[29] A. Hermoso-Carazo and J. Linares-Pérez, "Nonlinear estimation applying an unscented transformation in systems with correlated uncertain observations," *Applied Mathematics and Computation*, vol. 217, no. 20, pp. 7998–8009, 2011.

[30] S. Kolås, B. A. Foss, and T. Schei, "Constrained nonlinear state estimation based on the UKF approach," *Computers & Chemical Engineering*, vol. 33, no. 8, pp. 1386–1401, 2009.

[31] F. Gustafsson and G. Hendeby, "Some relation between extended and unscented Kalman filters," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 545–555, 2012.

[32] C. McGillem, J. Aunon, and D. Childers, "Signal processing in evoked potential research: applications of filtering and pattern recognition.," *Critical reviews in bioengineering*, vol. 6, no. 3, pp. 225–265, 1980.

[33] F. Daum, "Nonlinear filters: Beyond the Kalman filter," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, pp. 57–69, 2004.

[34] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.

[35] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *European conference on computer vision*, pp. 343–356, Springer, 1996.

[36] J. H. Halton, "Sequential monte carlo," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 58, pp. 57–78, Cambridge University Press, 1962.

[37] K. Kanazawa, D. Koller, and S. Russell, "Stochastic simulation algorithms for dynamic probabilistic networks," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 346–351, Morgan Kaufmann Publishers Inc., 1995.

[38] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.

[39] J. S. Liu and R. Chen, "Sequential monte carlo methods for dynamic systems," *Journal of the American statistical association*, vol. 93, no. 443, pp. 1032–1044, 1998.

[40] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 107–113, IET, 1993.

[41] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE signal processing magazine*, vol. 20, no. 5, pp. 19–38, 2003.

[42] P. J. Van Leeuwen, "Particle filtering in geophysical systems," *Monthly Weather Review*, vol. 137, no. 12, pp. 4089–4114, 2009.

[43] T. Adali and S. Haykin, *Adaptive signal processing: next generation solutions*, vol. 55. John Wiley & Sons, 2010.

[44] V. Cevher, A. C. Sankaranarayanan, J. H. McClellan, and R. Chellappa, "Target tracking using a joint acoustic video system," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 715–727, 2007.

[45] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *Journal of the American statistical association*, vol. 94, no. 446, pp. 590–599, 1999.

[46] C. K. Wikle and L. M. Berliner, "A Bayesian tutorial for data assimilation," *Physica D: Nonlinear Phenomena*, vol. 230, no. 1, pp. 1–16, 2007.

[47] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputations," *Journal of the american statistical association*, vol. 90, no. 430, pp. 567–576, 1995.

[48] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, 2002.

[49] J. Carpenter, P. Clifford, and P. Fearnhead, "Improved particle filter for nonlinear problems," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 146, no. 1, pp. 2–7, 1999.

[50] E. BøLviken, P. J. Acklam, N. Christophersen, and J.-M. StøRdal, "Monte carlo filters for non-linear state estimation," *Automatica*, vol. 37, no. 2, pp. 177–183, 2001.

[51] P. J. van Leeuwen, "Nonlinear data assimilation in geosciences: an extremely efficient particle filter," *Quarterly Journal of the Royal Meteorological Society*, vol. 136, no. 653, pp. 1991–1999, 2010.

[52] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[53] E. N. Lorenz, "Predictability: A problem partly solved," in *Proc. Seminar on predictability*, vol. 1, 1996.

[54] S. Nakano, G. Ueno, and T. Higuchi, "Merging particle filter for sequential data assimilation," *Nonlinear Processes in Geophysics*, vol. 14, no. 4, pp. 395–408, 2007.

[55] A. K. Singh, P. Date, and S. Bhaumik, "New algorithm for continuous-discrete filtering with randomly delayed measurements," *IET Control Theory & Applications*, vol. 10, no. 17, pp. 2298–2305, 2016.

[56] S. Allahyani and P. Date, "A minimum variance filter for discrete time linear systems with parametric uncertainty," in *MED'16: The 24th Mediterranean Conference on Control and Automation*, pp. 159–163, Mediterranean Control Association, 2016.

[57] S. Allahyani and P. Date, "A new approximate minimum variance filter for discrete time linear systems with randomly delayed observations and additive-multiplicative noise," *Digital Signal Processing, Submitted.*

[58] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[59] E. Gershon, U. Shaked, and I. Yaesh, "H_∞-control and filtering of discrete-time stochastic systems with multiplicative noise," *Automatica*, vol. 37, no. 3, pp. 409–417, 2001.

[60] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, vol. 23. Springer, 1992.

[61] E. Todorov and M. I. Jordan, "A minimal intervention principle for coordinated movement," in *Advances in neural information processing systems*, pp. 27–34, 2003.

[62] R. R. Mohler, *Bilinear control processes: with applications to engineering, ecology and medicine.* Academic Press, Inc., 1973.

[63] O. Besson and F. Castanié, "On estimating the frequency of a sinusoid in autoregressive multiplicative noise," *Signal processing*, vol. 30, no. 1, pp. 65–83, 1993.

[64] S. Kajita and F. Itakura, "Robust speech feature extraction using sbcor analysis," in *1995 International Conference on Acoustics, Speech, and Signal Processing, 1995. ICASSP-95.*, vol. 1, pp. 421–424, IEEE, 1995.

[65] R. T. Frankot and R. Chellappa, "Lognormal random-field models and their applications to radar image synthesis," *IEEE Transactions on Geoscience and Remote Sensing*, no. 2, pp. 195–207, 1987.

[66] R. F. Dwyer, "Fourth-order spectra of gaussian amplitude-modulated sinusoids," *The Journal of the Acoustical Society of America*, vol. 90, no. 2, pp. 918–926, 1991.

[67] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Transactions on pattern analysis and machine intelligence*, no. 2, pp. 157–166, 1982.

[68] D. Makrakis and P. Mathiopoulos, "Prediction/cancellation techniques for fading broadcasting channels. i. PSK signals," *IEEE Transactions on Broadcasting*, vol. 36, no. 2, pp. 146–155, 1990.

[69] P. McFadden, "Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration," *Journal of vibration, acoustics, stress, and reliability in design*, vol. 108, no. 2, pp. 165–170, 1986.

[70] R. Randall, "Cepstrum analysis and gearbox fault-diagnosis," *Maintenance Management International*, vol. 3, no. 3, pp. 183–208, 1982.

[71] G. Calafiore, "Reliable localization using set-valued nonlinear filters," *IEEE Transactions on systems, man, and cybernetics-part A: systems and humans*, vol. 35, no. 2, pp. 189–197, 2005.

[72] X. Kai, C. Wei, and L. Liu, "Robust extended Kalman filtering for nonlinear systems with stochastic uncertainties," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 2, pp. 399–405, 2010.

[73] W. Li, E. Todorov, and R. E. Skelton, "Estimation and control of systems with multiplicative noise via linear matrix inequalities," in *Proceedings of the American Control Conference*, pp. 1811–1816, IEEE, 2005.

[74] F. Carravetta, A. Germani, and N. Raimondi, "Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise," *IEEE Transactions on Automatic Control*, vol. 42, no. 8, pp. 1106–1126, 1997.

[75] O. L. Costa and G. R. Benites, "Linear minimum mean square filter for discrete-time linear systems with markov jumps and multiplicative noises," *Automatica*, vol. 47, no. 3, pp. 466–476, 2011.

[76] F. Wang and V. Balakrishnan, "Robust Kalman filters for linear time-varying systems with stochastic parametric uncertainties," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 803–813, 2002.

[77] K. Ponomareva and P. Date, "An exact minimum variance filter for a class of discrete time systems with random parameter perturbations," *Applied Mathematical Modelling*, vol. 38, no. 9, pp. 2422–2434, 2014.

[78] B. Chow and W. P. Birkemeier, "A new structure of recursive estimator," *IEEE Transactions on Automatic Control*, vol. 34, no. 5, pp. 586–572, 1989.

[79] B. Chow and W. Birkemeier, "A new recursive filter for systems with multiplicative noise," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1430–1435, 1990.

[80] F. Yang, Z. Wang, and Y. Hung, "Robust Kalman filtering for discrete time-varying uncertain systems with multiplicative noises," *IEEE Transactions on Automatic Control*, vol. 47, no. 7, pp. 1179–1183, 2002.

[81] J. Ishihara, M. Terra, and J. Cerri, "Optimal robust filtering for systems subject to uncertainties," *Automatica*, vol. 52, pp. 111–117, 2015.

[82] O. Vasicek, "An equilibrium characterization of the term structure," *Journal of financial economics*, vol. 5, no. 2, pp. 177–188, 1977.

[83] P. Date and R. Bustreo, "Value-at-risk for fixed-income portfolios: a Kalman filtering approach," *IMA Journal of Management Mathematics*, vol. 27, no. 4, pp. 557–573, 2016.

[84] P. Date and C. Wang, "Linear Gaussian affine term structure models with unobservable factors: Calibration and yield forecasting," *European Journal of Operational Research*, vol. 195, no. 1, pp. 156–166, 2009.

[85] S. H. Babbs and K. B. Nowman, "Kalman filtering of generalized vasicek term structure models," *Journal of Financial and Quantitative Analysis*, vol. 34, no. 1, pp. 115–130, 1999.

[86] A. C. Harvey, *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.

[87] S. Sun, L. Xie, W. Xiao, and Y. C. Soh, "Optimal linear estimation for systems with multiple packet dropouts," *Automatica*, vol. 44, no. 5, pp. 1333–1342, 2008.

[88] R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Least-squares linear estimators using measurements transmitted by different sen-

sors with packet dropouts," *Digital Signal Processing*, vol. 22, no. 6, pp. 1118–1125, 2012.

[89] S. L. Sun, "Optimal estimators for systems with finite consecutive packet dropouts," *IEEE Signal Processing Letters*, vol. 16, no. 7, pp. 557–560, 2009.

[90] R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Recursive least-squares quadratic smoothing from measurements with packet dropouts," *Signal Processing*, vol. 92, no. 4, pp. 931–938, 2012.

[91] S. Sun, L. Xie, W. Xiao, and N. Xiao, "Optimal filtering for systems with multiple packet dropouts," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 55, no. 7, pp. 695–699, 2008.

[92] N. Nahi, "Optimal recursive estimation with uncertain observation," *IEEE Transactions on Information Theory*, vol. 15, no. 4, pp. 457–462, 1969.

[93] S. Nakamori, R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Linear recursive discrete-time estimators using covariance information under uncertain observations," *Signal processing*, vol. 83, no. 7, pp. 1553–1559, 2003.

[94] Z. Wang, F. Yang, D. W. Ho, and X. Liu, "Robust finite-horizon filtering for stochastic systems with missing measurements," *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 437–440, 2005.

[95] Z. Wang, F. Yang, D. W. Ho, and X. Liu, "Robust H_∞ filtering for stochastic time-delay systems with missing measurements," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2579–2587, 2006.

[96] Z. Wang, D. W. Ho, and X. Liu, "Variance-constrained filtering for uncertain stochastic systems with missing measurements," *IEEE Transactions on Automatic control*, vol. 48, no. 7, pp. 1254–1258, 2003.

[97] H. Zhang, G. Feng, and C. Han, "Linear estimation for random delay systems," *Systems & Control Letters*, vol. 60, no. 7, pp. 450–459, 2011.

[98] X. Lu, L. Xie, H. Zhang, and W. Wang, "Robust Kalman filtering for discrete-time systems with measurement delay," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 6, pp. 522–526, 2007.

[99] S. Sun, "Linear minimum variance estimators for systems with bounded random measurement delays and packet dropouts," *Signal Processing*, vol. 89, no. 7, pp. 1457–1466, 2009.

[100] C. Han, H. Zhang, and M. Fu, "Optimal filtering for networked systems with Markovian communication delays," *Automatica*, vol. 49, no. 10, pp. 3097–3104, 2013.

[101] F. O. Hounkpevi and E. E. Yaz, "Minimum variance generalized state estimators for multiple sensors with different delay rates," *Signal Processing*, vol. 87, no. 4, pp. 602–613, 2007.

[102] A. RAY, L. LIOU, and J. SHEN, "State estimation using randomly delayed measurements," *Journal of dynamic systems, measurement, and control*, vol. 115, no. 1, pp. 19–26, 1993.

[103] Z. Wang and D. W. Ho, "Robust filtering under randomly varying sensor delay with variance constraints," in *European Control Conference (ECC), 2003*, pp. 3204–3209, IEEE, 2003.

[104] E. Yaz and A. Ray, "Linear unbiased state estimation under randomly varying bounded sensor delay," *Applied Mathematics Letters*, vol. 11, no. 4, pp. 27–32, 1998.

[105] A. Hermoso-Carazo and J. Linares-Pérez, "Linear and quadratic least-squares estimation using measurements with correlated one-step random delay," *Digital Signal Processing*, vol. 18, no. 3, pp. 450–464, 2008.

[106] S. Nakamori, R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Recursive estimators of signals from measurements with stochastic delays using covariance information," *Applied Mathematics and Computation*, vol. 162, no. 1, pp. 65–79, 2005.

[107] S. Sun and W. Xiao, "Optimal linear estimators for systems with multiple random measurement delays and packet dropouts," *International Journal of Systems Science*, vol. 44, no. 2, pp. 358–370, 2013.

[108] M. Choi, J. Choi, and W. Chung, "State estimation with delayed measurements incorporating time-delay uncertainty," *Control Theory & Applications, IET*, vol. 6, no. 15, pp. 2351–2361, 2012.

[109] R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Optimal state estimation for networked systems with random parameter matrices, correlated noises and delayed measurements," *International Journal of General Systems*, vol. 44, no. 2, pp. 142–154, 2015.

[110] D. Chen, Y. Yu, L. Xu, and X. Liu, "Kalman filtering for discrete stochastic systems with multiplicative noises and random two-step sensor delays," *Discrete Dynamics in Nature and Society*, vol. 2015, 2015.

[111] R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Fusion estimation using measured outputs with random parameter matrices subject to random delays and packet dropouts," *Signal Processing*, vol. 127, pp. 12–23, 2016.

[112] J. Hu, Z. Wang, B. Shen, and H. Gao, "Quantised recursive filtering for a class of nonlinear systems with multiplicative noises and missing measurements," *International Journal of Control*, vol. 86, no. 4, pp. 650–663, 2013.

[113] J. Ma and S. Sun, "Optimal linear estimators for multi-sensor stochastic uncertain systems with packet losses of both sides," *Digital Signal Processing*, vol. 37, pp. 24–34, 2015.

[114] J. Ma and S. Sun, "Centralized fusion estimators for multi-sensor systems with multiplicative noises and missing measurements," *Journal of Networks*, vol. 7, no. 10, pp. 1538–1545, 2012.

[115] H. Rezaei, R. M. Esfanjani, and M. H. Sedaaghi, "Improved robust finite-horizon Kalman filtering for uncertain networked time-varying systems," *Information Sciences*, vol. 293, pp. 263–274, 2015.

[116] S. Sun, T. Tian, and L. Honglei, "State estimators for systems with random parameter matrices, stochastic nonlinearities, fading measurements and correlated noises," *Information Sciences*, vol. 397, pp. 118–136, 2017.

[117] H. Qian, Z. Qiu, and Y. Wu, "Robust extended Kalman filtering for nonlinear stochastic systems with random sensor delays, packet dropouts and correlated noises," *Aerospace Science and Technology*, vol. 66, pp. 249–261, 2017.

[118] S. Wang, H. Fang, and X. Tian, "Minimum variance estimation for linear uncertain systems with one-step correlated noises and incomplete measurements," *Digital Signal Processing*, vol. 49, pp. 126–136, 2016.

[119] S. Wang, H. Fang, and X. Tian, "Recursive estimation for nonlinear stochastic systems with multi-step transmission delays, multiple packet dropouts and correlated noises," *Signal Processing*, vol. 115, pp. 164–175, 2015.

[120] J. Hu, Z. Wang, and H. Gao, "Recursive filtering with random parameter matrices, multiple fading measurements and correlated noises," *Automatica*, vol. 49, no. 11, pp. 3440–3448, 2013.

[121] X. Wang, Y. Liang, Q. Pan, and C. Zhao, "Gaussian filter for nonlinear systems with one-step randomly delayed measurements," *Automatica*, vol. 49, no. 4, pp. 976–986, 2013.

[122] A. Hermoso-Carazo and J. Linares-Pérez, "Unscented filtering algorithm using two-step randomly delayed observations in nonlinear systems," *Applied Mathematical Modelling*, vol. 33, no. 9, pp. 3705–3717, 2009.

[123] J. Hu, Z. Wang, H. Gao, and L. K. Stergioulas, "Extended Kalman filtering with stochastic nonlinearities and multiple missing measurements," *Automatica*, vol. 48, no. 9, pp. 2007–2015, 2012.

[124] S. Allahyani and P. Date, "A minimum variance filter for continuous discrete systems with additive-multiplicative noise," in *24th European Signal Processing Conference (EUSIPCO), 2016*, pp. 2330–2334, IEEE, 2016.

[125] S. Allahyani and P. Date, "A new approximate minimum variance filter for continuous-discrete linear systems with randomly delayed observations and additive-multiplicative noise," *Digital Signal Processing, Submitted.*

[126] A. H. Jazwinski, *Stochastic processes and filtering theory.* Courier Corporation, 2007.

[127] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations.* Berlin, Germany: Springer, 1999.

[128] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software.* John Wiley & Sons, 2004.

[129] K. J. Åström, *Introduction to stochastic control theory.* Courier Corporation, 2012.

[130] S. Blackman and R. Popoli, "Design and analysis of modern tracking systems," *Norwood, MA: Artech House, 1999.*, 1999.

[131] B. Oksendal, *Stochastic differential equations: an introduction with applications.* Springer Science & Business Media, 2013.

[132] P. S. Maybeck, *Stochastic models, estimation, and control*, vol. 3. Academic press, 1982.

[133] A. Doucet, N. de Freitas, and N. Gordon, "Sequential monte carlo methods in practice," *Statistics*, 2001.

[134] S. Särkkä, "On unscented Kalman filtering for state estimation of continuous-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1631–1641, 2007.

[135] I. Arasaratnam, S. Haykin, and T. R. Hurd, "Cubature Kalman filtering for continuous-discrete systems: theory and simulations," *Signal Processing, IEEE Transactions on*, vol. 58, no. 10, pp. 4977–4993, 2010.

[136] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, 2009.

[137] S. Särkkä and A. Solin, “On continuous-discrete cubature Kalman filtering,” *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 1221–1226, 2012.

[138] S. Särkkä and J. Sarmavuori, “Gaussian filtering and smoothing for continuous-discrete dynamic systems,” *Signal Processing*, vol. 93, no. 2, pp. 500–510, 2013.

[139] F. E. Daum, “Exact finite dimensional nonlinear filters for continuous time processes with discrete time measurements,” in *The 23rd IEEE Conference on Decision and Control, 1984*, vol. 23, pp. 16–22, IEEE, 1984.

[140] M. Mallick, M. Morelande, and L. Mihaylova, “Continuous-discrete filtering using EKF, UKF, and PF,” in *2012 15th International Conference on Information Fusion (FUSION)*, pp. 1087–1094, IEEE, 2012.

[141] S. Xu and T. Chen, “Reduced-order H_∞ filtering for stochastic systems,” *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 2998–3007, 2002.

[142] J. C. Jiménez and T. Ozaki, “Linear estimation of continuous-discrete linear state space models with multiplicative noise,” *Systems & control letters*, vol. 47, no. 2, pp. 91–101, 2002.

[143] J. Jimenez and T. Ozaki, “Local linearization filters for non-linear continuous-discrete state space models with multiplicative noise,” *International Journal of Control*, vol. 76, no. 12, pp. 1159–1170, 2003.

[144] P. Rajasekaran, N. Satyanarayana, and M. Srinath, “Optimum linear estimation of stochastic signals in the presence of multiplicative noise,” *IEEE Transactions on Aerospace and Electronic Systems*, no. 3, pp. 462–468, 1971.

[145] J. Jimenez and T. Ozaki, “Linear estimation of continuous-discrete linear state space models with multiplicative noise,” *Systems & control letters*, vol. 47, no. 2, pp. 91–101, 2002.

[146] D. Hinrichsen and A. J. Pritchard, “Stochastic H_∞,” *SIAM Journal on Control and Optimization*, vol. 36, no. 5, pp. 1504–1538, 1998.

[147] P. Shi, E.-K. Boukas, and R. K. Agarwal, "Kalman filtering for continuous-time uncertain systems with markovian jumping parameters," *IEEE Transactions on Automatic Control*, vol. 44, no. 8, pp. 1592–1597, 1999.

[148] J. C. Cox, J. E. Ingersoll, and S. A. Ross, "A theory of the term structure of interest rates," *Econometrica*, vol. 53, no. 2, pp. 385–407, 1985.

[149] H. Zhang, X. Lu, and D. Cheng, "Optimal estimation for continuous-time systems with delayed measurements," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 823–827, 2006.

[150] V. Kolmanovskii, T. Maizenberg, and J.-P. Richard, "Mean square stability of difference equations with a stochastic delay," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 52, no. 3, pp. 795–804, 2003.

[151] C. E. De Souza, R. M. Palhares, and P. L. Peres, "Robust H_∞ filtering for uncertain linear systems with multiple time-varying state delays: an lmi approach," in *Proceedings of the 38th IEEE Conference on Decision and Control, 1999.*

[152] H. Zhang, X. Lu, W. Zhang, and W. Wang, "Kalman filtering for linear time-delayed continuous-time systems with stochastic multiplicative noises," *International Journal of Control, Automation, and Systems*, vol. 5, no. 4, pp. 355–363, 2007.

[153] S. Xu, T. Chen, and J. Lam, "Robust H_∞ filtering for uncertain markovian jump systems with mode-dependent time delays," *IEEE Transactions on Automatic control*, vol. 48, no. 5, pp. 900–907, 2003.

[154] C. E. de Souza, R. M. Palhares, and P. D. Peres, "Robust H_∞ filter design for uncertain linear systems with multiple time-varying state delays," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 569–576, 2001.

[155] H. Zhang and L. Xie, *Control and estimation of systems with input/output delays.* Springer Publishing Company, Incorporated, 2007.

[156] W.-W. Che, Y.-P. Li, and Y.-L. Wang, "H_∞ tracking control for NCS with packet losses in multiple channels case," *International Journal of Innovative Computing Information and Control*, vol. 2011, no. 7, pp. 6507–6522, 2011.

[157] J. Nilsson, B. Bernhardsson, and B. Wittenmark, "Stochastic analysis and control of real-time systems with random time delays," *Automatica*, vol. 34, no. 1, pp. 57–64, 1998.

[158] H. Dong, Z. Wang, and H. Gao, "Robust H_∞ filtering for a class of nonlinear networked systems with multiple stochastic communication delays and packet dropouts," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 1957–1966, 2010.

[159] A. Hermoso-Carazo and J. Linares-Pérez, "Extended and unscented filtering algorithms using one-step randomly delayed observations," *Applied Mathematics and Computation*, vol. 190, no. 2, pp. 1375–1393, 2007.

[160] C.-L. Su and C.-N. Lu, "Interconnected network state estimation using randomly delayed measurements," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 870–878, 2001.

[161] Z. Wang, J. Lam, and X. Liu, "Filtering for a class of nonlinear discrete-time stochastic systems with state delays," *Journal of Computational and Applied Mathematics*, vol. 201, no. 1, pp. 153–163, 2007.

[162] P. Date and S. Allahyani, "An approximate minimum variance filter for nonlinear systems with randomly delayed observations," in *25th European Signal Processing Conference (EUSIPCO), 2017*, pp. 1639–1643, IEEE, 2017.