LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
LUT School of Engineering Science
Department of Mathematics and Physics

Dominique Ingabe Kalisa

Initialization of Continuous Nonlinear Models Using Extended Kalman Filter

Supervisors: Professor Heikki Haario
D.Sc. (Tech.) Isambi Sailon Mbalawata

Examiners: Professor Heikki Haario
D.Sc. (Tech.) Marko Laine

ABSTRACT

Lappeenranta University of Technology

LUT School of Engineering Science

Department of Mathematics and Physics

Dominique Ingabe Kalisa

Initialization of Continuous Nonlinear Models Using Extended Kalman Filter

Master's thesis

2015

52 pages, 12 figures, 4 tables

Supervisors: Professor Heikki Haario

D.Sc. (Tech.) Isambi Sailon Mbalawata

Examiners: Professor Heikki Haario

D.Sc. (Tech.) Marko Laine

Keywords: state-space models, Kalman filter, diffuse initial conditions, Markov Chain Monte Carlo (MCMC), parameter estimation

The two main objectives of Bayesian inference are to estimate parameters and states. In this thesis, we are interested in how this can be done in the framework of state-space models when there is a complete or partial lack of knowledge of the initial state of a continuous nonlinear dynamical system. In literature, similar problems have been referred to as diffuse initialization problems. This is achieved first by extending the previously developed diffuse initialization Kalman filtering techniques for discrete systems to continuous systems. The second objective is to estimate parameters using MCMC methods with a likelihood function obtained from the diffuse filtering. These methods are tried on the data collected from the 1995 Ebola outbreak in Kikwit, DRC in order to estimate the parameters of the system.

Acknowledgements

First and foremost, I would like to express my gratitude to the Department of Mathematics for the opportunity to further pursue my studies in a collaborative and learning-friendly environment as well as for the financial support provided throughout the course of my studies.

My sincere thanks to Professor Heikki Haario for his guidance and support towards the completion of this thesis. I would also like to extend my heartfelt gratitude to my co-supervisor Dr. Isambi S. Mbalawata for his insightful and continuous assistance.

To my friends and colleagues, thank you for making this journey memorable.

Last but not least, I would like to thank my family for their love and support. My deepest love and appreciation are addressed to my father Professor Daniel Kalisa, whose wise counsel and unfailing support, have brought me this far.

Lappeenranta, May 7th, 2015

Dominique Ingabe Kalisa

Contents

List of Symbols and Abbreviations

KF	Kalman Filter
MCMC	Markov Chain Monte Carlo
DRC	Democratic Republic of Congo
SSM	State-Space Model
EKF	Extended Kalman Filter
DKF	Diffuse Kalman Filter
ODE	Ordinary Differential Equation
CD-EKF	Continuous-Discrete Extended Kalman Filter
MC	Monte carlo
MAP	Maximum A Posteriori
MLE	Maximum Likelihood Estimator
RWM	Random Walk Metropolis
AM	Adaptive Metropolis
SIR	Susceptible-Infected-Recovered
EHF	Ebola Hemorrhagic Fever
SEIR	Susceptible-Exposed-Infected-Recovered
CDC	Center for Disease Control
WHO	World Health Organisation
KFS	Kalman Filter and Smoother
Prob.	Probability
1D	One dimensional
2D	Two dimensional
KFS	Kalman Filter-Smoother

1 INTRODUCTION

Natural phenomena are usually modeled using differential equations that describe its time evolution using variables of interest that adequately represent it. The mathematical framework within which the evolution of the phenomenon is studied, is called a *dynamical system*. Dynamic systems theory finds its origin in control theory, an interdisciplinary branch of engineering and mathematics, that studies the behavior of dynamic systems. Through measurement devices, researchers/engineers can *observe* the states of the system for given lengths of time. In the state-space approach, mathematical models and observations from the system can be combined to estimate the states of the system at any given time. This combination can be seen in the formulation of a state-space model, which is made of a *state equation* and a *measurement equation* (Hamilton, 1994a).

Considering that no mathematical model or measurement device can provide perfectly reliable information about a system, there are state estimation methods that take into account the noise corrupted model and measurements. For linear and Gaussian systems in particular, the famous Kalman filter developed by R. E. Kalman (1960), is a data processing algorithm which was found to produce optimal estimates while using all available information provided to it regardless of its accuracy (Maybeck, 1979) . Suboptimal extensions of the KF-type such as the extended Kalman filter or the unscented Kalman filter, have been developed to handle nonlinear and continuous systems that represent physical systems more accurately.

In addition, the recursive nature of the algorithm enables to estimate the state of the system using its most recent estimate. This is a fundamental property in the derivation of the different Kalman filter algorithm types. Therefore, the execution of the filtering algorithms is straightforward provided there is enough prior knowledge of the system to get the filter started. Traditionally, insufficient or lack of knowledge of the initial conditions is treated by assigning a rather large covariance to the initial state (Harvey and Phillips, 1979; Schweppe, 1973). This approach can be numerically inefficient.

The concept of a filter that could be initialized by accounting for total or partial lack of knowledge of the initial state was introduced in a series of papers by Ansley and Kohn (1985, 1989) and Kohn and Ansley (1986). De Jong (1991) further developed and presented an easier algorithm to implement in which the states and innovation vectors are augmented by matrices indicating the diffuseness of the initial distribution. The extra recursions introduced by the augmentation vanish when the diffuse

vector is identified.

Also based on the ideas introduced in Ansley and Kohn (1985) , Koopman (1997) and Koopman and Durbin (2003) treat the same issue with a different approach. The initial covariance is decomposed into a diffuse part and a proper part, each part having their own update equations until the effects of the diffuseness vanish. In both cases, after the diffuse effects disappear, both algorithms fall back to the regular Kalman filter.

The work of this thesis is based on the work of Koopman and Durbin (2003) and aims at using their method to initialize other variants of the Kalman Filter for continuous and/or nonlinear dynamic systems. The second objective of the thesis is to use the likelihood function determined in the previous step for parameter estimation and uncertainty analysis using Markov chain Monte Carlo Methods

This thesis is organized as follows. The next section consists of a review of filtering and smoothing methods first for linear and non linear models for discrete dynamic systems and then for continuous systems. An emphasis is put on the diffuse initialisation of each these methods. Section 3 is a brief introduction to Markov chain Monte Carlo methods for parameter estimation and finally section 4 presents an application of the diffuse filtering in parameter estimation using the filtering likelihood function in Markov Chain Monte Carlo methods. Conclusions are given in section 5.

2 State Space Models

State space models (SSM) are a set of two probabilistic equations often used in the analysis of dynamical systems. They allow to infer the conditional distribution of a latent variable called the state vector given observed aspects of the system that are either relevant to the problem or accessible. A discrete linear SSM is described as follows (Durbin and Koopman, 2001):

$$x_{t+1} = T_t x_t + R_t \epsilon_t, \quad \epsilon_t \sim N(0, Q_t) \tag{2.1a}$$

$$y_t = Z_t x_t + \zeta_t, \quad \zeta_t \sim N(0, H_t), \quad t = 1, ..., n. \tag{2.1b}$$

Equations 2.1a and 2.1b are respectively called the evolution and observation model. The evolution model describes the propagation of the state in time where as the observation model relates the observations to the state. The terms ϵ_t and ζ_t are respectively the process and measurement noise which are zero mean Gaussian distributed with covariance, Q_t and H_t. The matrices T_t, Z_t, R_t, Q_t and H_t are known and can be time dependent or not. The SSM framework is able to represent linear and nonlinear systems. The first part of this section derives statistical tools for linear models, nonlinear SSMs are introduced later.

Three different problems can arise in dynamical state estimation: smoothing, filtering and forecasting (Särkkä, 2013).

- Smoothing : The state of the system at time is estimated using the entire stack of available information
- Filtering : The state of the system at time is estimated "on-line" as new measurements are obtained.
- Forecasting : The state of the system is predicted k steps ahead using previous measurements.

In 1960, Rudolph E. Kalman developed the Kalman filter (KF) (Kalman, 1960), a computationally efficient algorithm for the estimation of discrete data linear SSM. Since then it has extensively been studied and applied to nonlinear state space models by linearizing the nonlinear evolution and observation models and incorporating them in the KF algorithm as first developed by R.E Kalman. This modified KF is

often referred to as the extended Kalman filter(EKF). There are several versions of the Kalman filter adapted to suit different real life situations but these are outside the scope of this work.

2.1 Kalman Filter

The purpose of the Kalman filter is to compute the conditional distribution of x_{t+1} given the observations $y_{1:t} = \{y_1, y_2, ..., y_t\}$ for $t = 1, .., n$. Its popularity lies in its properties as an optimal and recursive algorithm. Thanks to the Markovian property of the evolution model, there is no need to store and process previous data when a new measurement is provided to the filter; the current state estimate is determined by the current measurement and the previous state estimate. This makes the algorithm recursive and efficient from a computational point of view. Optimality is obtained by minimizing the mean squared error under the assumptions of model linearity and Gaussian white measurement and process noise (Maybeck, 1979). Given the normality assumption of the distributions in the SSM, we can write the conditional distribution of x_{t+1} in terms of its first two moments as

$$a_{t+1} = E(x_{t+1}|y_{1:t}), \tag{2.2a}$$

$$P_{t+1} = Var(x_{t+1}|y_{1:t}). \tag{2.2b}$$

Below are the set of equations that constitute the KF algorithm for (2.1a) and (2.1b). The derivations can be found in (Durbin and Koopman, 2001).

Algorithm 1 Kalman filter

Initialize Kalman filter with (a_1, P_1)

for all $t = 1, ..., n$ **do**

$v_t = y_t - Z_t a_t$

$F_t = Z_t P_t Z_t' + H_t$

$K_t = P_t Z_t' F_t^{-1}$

$a_{t|t} = a_t + K_t v_t$

$a_{t+1} = T_t a_{t|t}$

$P_{t|t} = P_t - K_t F_t K_t'$

$P_{t+1} = T_t P_{t|t} T_t' + R_t Q_t R_t'$

end for

The previously predicted state estimate a_t can be written in the form of equation 2.2a as $a_t = E(x_t|y_{1:t-1})$. It is "filtered" via the new information in the form of v_t, the innovation vector containing new information provided by the most recent observation. The term v_t is weighted by the *Kalman gain*, which can be intuitively explained as the measure of trust granted to either state estimates or new measurements. The filtered estimate $a_{t|t}$ is then predicted to a_{t+1}. The state variance P_{t+1} is similarly obtained.

To start the filter, (a_1, P_1) are assumed to be given or known although it is often the case that some or all initial states are unknown thus rendering the KF unusable (Ansley and Kohn, 1989). It is however a common practice to initialize the KF with guessed initial conditions picked from a range of reasonable values, hoping that the filter will "forget" the random guess and converge towards the solution rapidly. While this is acceptable when there is enough data, it is a luxury that cannot be afforded with small data sets. To illustrate this, let's consider a time series (see Koopman (1997), Harvey (1989)) with a time-varying trend μ_t and a time-varying β_t given as below for $t = 1, \cdots, n$,

$$
\begin{aligned}
x_t &= \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}}_{T} x_t + \begin{pmatrix} \sigma_\mu & 0 \\ 0 & \sigma_\beta \end{pmatrix} \epsilon_t, \qquad (2.3) \\
y_t &= \underbrace{\begin{pmatrix} 1 & 0 \end{pmatrix}}_{Z} x_t + \begin{pmatrix} \sigma_y & 0 \end{pmatrix} \epsilon_t, \quad \epsilon_t \sim N(0, I_2)
\end{aligned}
$$

Not knowing the initial states of the above system, we simply take three different sets of guesses and observe the impact of each on the estimation. The SSM 2.3 is simulated using a fixed set initial values $[\mu_0, \beta_0] = [0.6, 0.95]$ regarded as the true initial state of the system. The trend is observed for 15 time periods, that is, $n = 15$. Figure 2.1 compares the three different system estimates obtained using different starting values. Two of the solutions (green and black dotted lines) start off far from the true trend and from each other, they only start to converge at $t = 10$. The third solution (magenta line) starts closer to the true trend but follows the others at $t = 10$ as well. We can thus see, that more than half the observations are used before the effect of each initial value dissipates.

In the next section, a variant of the Kalman filter that accommodates the lack of knowledge of initial conditions is introduced.

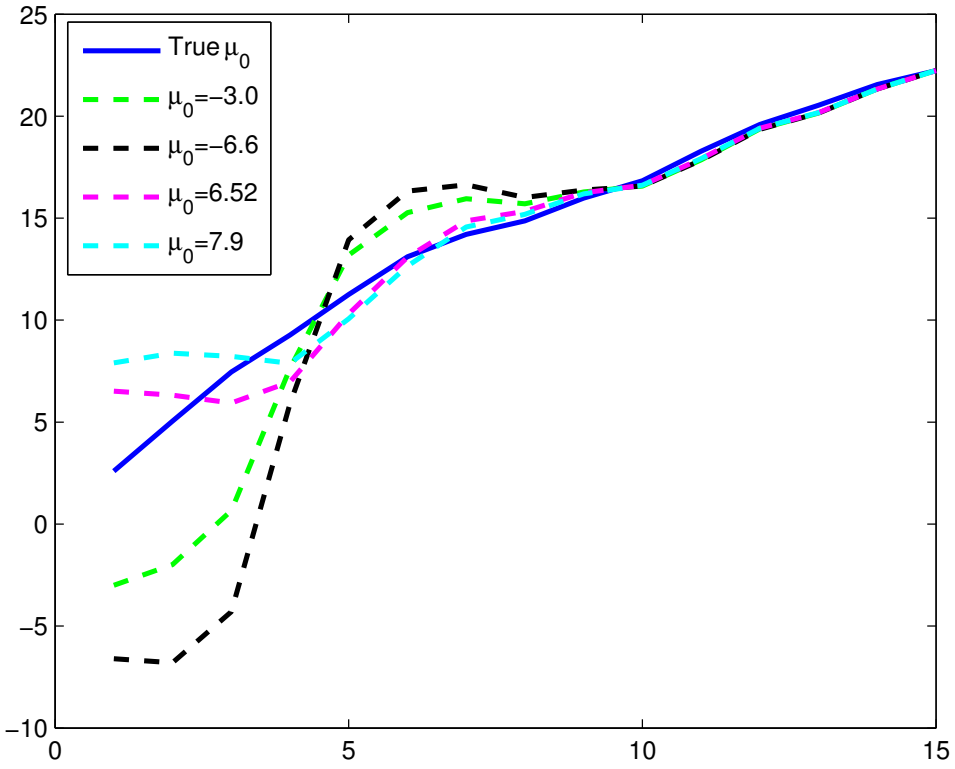

Figure 1: Initialisation of the Kalman filter with 3 randomly guessed sets of initial values for the first component μ_t. The filtered estimates in each case are compared to the true linear trend μ.

2.1.1 Kalman Filter with Unknown Initial Conditions

Mathematical models that describe a real life problem always carry a certain level of uncertainty. This uncertainty has a number of causes: simplification of the phenomenon for modeling purposes, use of numerical methods to approximate solutions to problems that are not analytically solvable or insufficient of knowledge of necessary model inputs such as initial conditions, control parameters, etc. Among these, initial values play a great role on the uncertainty of the parameters. Indeed, small perturbations in the initial values can propagate into huge errors down the road. Modelers have dealt with the issue of unknown initial values differently, such as estimating the initial state of the system along with the parameters (Bowong and Kurths, 2010) or considering the initial values as random variables and assigning them a certain distribution (Kegan and West, 2005; Omar and Hasan, 2012).

Although one of the most important tools for dynamical state estimation, the ordinary Kalman filter requires the algorithm to be initialized by known initial values. Besides the idea of letting the filter forget the random initial guesses provided to it, this issue has been circumvented by initializing the KF with a large covariance matrix, a method referred to in literature as the *big-K* method, that represent the lack of knowledge surrounding the initial state of the system, but in practice this method can lead to large rounding errors. As an alternative to the big-K method, the information filter as described in Anderson and Moore (1979) can also used in this case. However, Kohn and Ansley (1984a) show that for a particular order of ARIMA(p, d, q) models (where $p + d < q + 1$), the information filter cannot handle unknown initial conditions.

Before going further, we will introduce the term *diffuse initial conditions*. A system has diffuse initial conditions if its initial states have an arbitrarily large covariance. The initials states can be completely or partially diffuse depending on the extent of our prior knowledge.

In Ansley and Kohn (1985), a rather complex modification to the Kalman filter was developed to analytically handle diffuse initial conditions and overcome the drawbacks mentioned above in the initialization of the Kalman filter. The initial state and covariance is defined as (Ansley and Kohn, 1985),

$$x_1 = a + A\delta + R_0\epsilon_0 \quad \delta \sim N(0, \kappa I_q), \quad \epsilon \sim N(0, Q_0), \tag{2.4a}$$

$$P_1 = \kappa P_\infty + P_*, \quad \kappa \to \infty \tag{2.4b}$$

The $m \times 1$ vector a_1 is regarded as the known or proper part of x_1 whereas the $q \times 1$ random vector δ is regarded as representing the diffuse part of x_1. The covariance matrix of x_1 is similarly split in two components.

This modified Kalman filter laid the ground for two distinct alternatives both with a similar postulation of the initial state and covariance: the *diffuse Kalman Filter* by De Jong (1991) and the *exact initial Kalman filter* by Koopman (1997) (hereafter referred to as the DKF and EIKF respectively). These two approaches differ in how they adapt the ordinary Kalman filter as given in Algorithm 1 to accommodate the diffuseness in the initial conditions.

In De Jong (1991), the state a_{t+1} and innovations v_t vectors are column-augmented by A_{t+1} and V_{t+1} respectively for an initial length of time $t = 1, 2, \cdots, d - 1 \leq n$. The matrices (A_{t+1}, a_{t+1}) and (V_{t+1}, v_{t+1}) are $m \times (q + 1)$ matrices, where m is the

number of states and q is the number of components in δ. The algorithm also includes an additional matrix recursion Q_t for likelihood evaluation. When at $t = d \leq n$, the upper block of Q_t becomes invertible, the DKF collapses to the ordinary KF. While the DKF offers the possibility of explicitly recovering the diffuse vector δ, the collapse is not automatic as in the exact initial Kalman filter and can in some cases lead to $d = n+1$, which would mean that the state and innovation vector have been augmented by n columns. Being more dependent on the number of rows computed, the performance of the KF is usually not significantly affected by this augmentation (Chu-Chun-Lin, 1991). However, the possibility of a non- collapse is inconvenient.

On the other hand, Koopman (1997) proceeds to develop the EIKF with the idea of a diffuse and non-diffuse part where the variance-covariance matrix P_t is written as in Equation (2.4b):

$$P_t = \kappa P_{\infty,t} + P_{*,t} + O(\kappa^{-1}), \tag{2.5}$$

such that P_∞ and P_* do not depend on κ. This formulation is extended on two other quantities of the KF, the innovation covariance F_t and the Kalman gain K_t, which become

$$F_t = \kappa F_{\infty,t} + F_{*,t} + O(\kappa^{-1}), \tag{2.6}$$

$$K_t = \kappa K_{\infty,t} + K_{*,t} + O(\kappa^{-1}) \tag{2.7}$$

The derivation of the EIKF is based on the power series expansion of F_t^{-1} in κ^{-1},

$$F_t^{-1} = [\kappa F_{\infty,t} + F_{*,t} + O(\kappa^{-1})]^{-1}, \tag{2.8}$$

$$= F_t^{(0)} + \kappa^{-1} F_t^{(1)} + \kappa^{-2} F_t^{(2)} + O(\kappa^{-3}) \tag{2.9}$$

when $\kappa \to \infty$. Equation (2.8) is later used in the Kalman gain expression in Equation (2.7). The expansion allows to re-write the Kalman filter algorithm in such a way that the filter recursions involving the terms P_t, F_t and K_t are computed for the diffuse and proper components separately and independently from κ. For more theoretical details on the derivations of the EIKF, the proofs can be found in Durbin and Koopman (2001).

For the diffuse initial states elements, given the matrix structure of A , $P_{\infty,1}$ is a diagonal matrix with q elements on the diagonal and the rest equal to zero. To each non-zero element of $P_{\infty,1}$ corresponds an element of a equal to zero. Similarly, $P_{*,1}$ is a diagonal matrix with $m-q$ non-zero elements on the diagonal and the rest equal to zero. A reformulation of the Kalman filter can be seen in Algorithm 2.

Algorithm 2 Exact Initial Kalman filter

Initialize the Diffuse Kalman filter with (a_1, A), (P_∞, P_*)

for all $t = 1...d \leq n$ **do**

if $F_{\infty,t}$ is nonsingular **then**

$v_t = y_t - Z_t a_t$

$F_{\infty,t} = Z_t P_{\infty,t} Z_t'$

$F_{*,t} = Z_t P_{*,t} Z_t' + H_t$

$K_{\infty,t} = P_{\infty,t} Z_t' F_{\infty,t}^{-1}$

$K_{*,t} = (P_{*,t} Z_t' - K_{\infty,t} F_{*,t}) F_{\infty,t}^{-1}$

$P_{\infty,t|t} = P_{\infty,t} - K_{\infty,t} F_{\infty,t} K_{\infty,t}'$

$P_{*,t|t} = P_{*,t} - K_{\infty,t} Z_t P_{*,t}' - K_{*,t} Z_t P_{\infty,t}'$

$P_{\infty,t+1} = T_t P_{\infty,t|t} T_t'$

$P_{*,t+1} = T_t P_{*,t|t} T_t' + R_t Q_t R_t'$

$a_{t|t} = a_t + K_{\infty,t} v_t$

$a_{t+1} = T_t a_{t|t}$

else

if $F_{\infty,t} = 0$ **then**

$v_t = y_t - Z_t a_t$

$F_{*,t} = Z_t P_{*,t} Z_t' + H_t$

$K_{*,t} = P_{*,t} Z_t' F_{*,t}^{-1}$

$P_{\infty,t|t} = P_{\infty,t}$

$P_{*,t|t} = P_{*,t} - K_{*,t} Z_t P_{*,t}'$

$P_{*,t+1} = T_t P_{*,t|t} T_t' + R_t Q_t R_t'$

$a_{t|t} = a_t + K_{*,t} v_t$

$a_{t+1} = T_t a_{t|t}$

end if

end if

end for

for all $t = d + 1...n$ **do**

Use the algorithm 1

end for

It is easily seen that when all initial states are known $P_\infty = 0$ and the usual KF can be applied. Otherwise, the exact initial Kalman filter of Koopman and Durbin runs first for an initial stretch $t = 1, ..., d$ and collapses automatically to the KF when the influence of κ dies out (Koopman, 1997), that is, when $P_\infty = 0$. Both the KF and the exact initial KF require the inversion of the matrix F_t. In univariate series, F_t is a scalar and there is no singularity problem. In a multivariate setting, singularity is a rare instance but can happen to $F_{\infty,t}$, the component of F_t associated with $P_{\infty,t}$. In such a case, (Durbin and Koopman, 2001) suggest to transform the multivariate observation series into a univariate series by adding one component after the other.

The approach developed by Koopman (1997) is chosen as the filtering algorithm to be used in this thesis for its transparent treatment of the diffuse initial conditions and its more straightforward conditions for the collapse of the exact initial Kalman filter to the ordinary Kalman filter.

Algorithm 2 is implemented in the local linear trend model 2.3 in subsection1. The EIKF runs for $t = 1, 2$ and the Kalman filter starts at $t = 3$ with a_3 and $P_3 = P_{*,3}$. Figure 2.1.1 shows that at $t = 3$ the EIKF solution immediately jump towards the true solution faster at $t = 3$ and goes on to closely follow the other solutions.

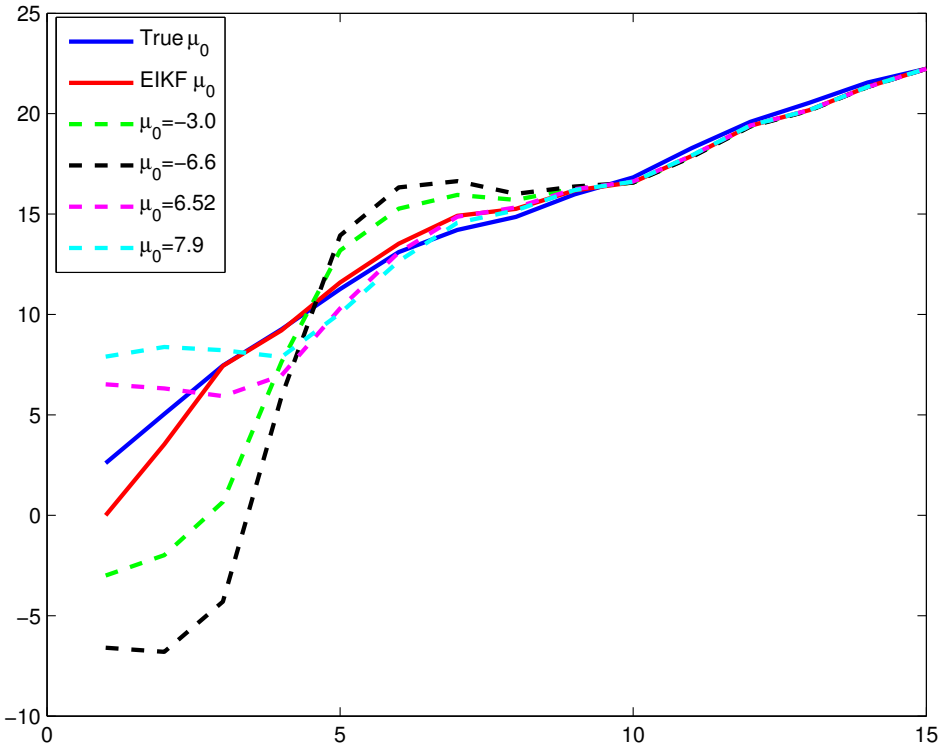

Figure 2: Exact initial Kalman filter compared(Red dotted line) to the ordinary Kalman filter with randomly guessed initial conditions.

2.1.2 Gaussian Likelihood Function

Given a data sample y of size n and a model :

$$x_{t+1} = T(x, \theta) + \epsilon, \quad \epsilon \sim N(0, Q)$$

$$y_t = Zx_t + \zeta, \quad \epsilon \sim N(0, H)$$

The likelihood function $l(y|\theta)$ is the probability of observing the measurements y given the unknown parameters θ.

$$l(y|\theta) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|F_t|}} exp\left(-\frac{1}{2}\sum_{i=1}^{n} v_t' F_t^{-1} v_t\right),$$

where the innovations v_t and their covariance matrix F_t are obtained via the Kalman filter.

In several cases, the natural logarithm of the likelihood function is more convenient to work with than the likelihood function itself. Indeed, the natural logarithm

increases monotonically and reaches its maximum at the same point the function itself reaches its own. Thus the Gaussian log-likelihood function is given by:

$$\log l(y|\theta) = -\frac{n}{2}log2\pi - \frac{1}{2}\sum_{t=1}^{n}\log|F_t| - \frac{1}{2}\sum_{t=1}^{n}v_t^{'}F_t^{-1}v_t, \tag{2.10}$$

Since the likelihood function depends on the innovation covariance matrix, it also modified to suit the changes made to the Kalman filter when there is a insufficient knowledge about initial conditions. The diffuse log-likelihood function, whose derivations will not be given here but can be found in (Koopman, 1997), is defined as

$$\log l_{\infty}(y|\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{d}w_t - \frac{1}{2}\sum_{t=d}^{n}\log|F_t| - \frac{1}{2}\sum_{t=d}^{n}v_t^{'}F_t^{-1}v_t, \tag{2.11}$$

where $w_i = \log|F_{\infty,t}|$ if $F_{\infty,t}$ is non-singular and $w_i = \log|F_{*,t}| + v_t^{'}F_{*,t}^{-1}v_t$ otherwise.

2.2 Smoothing

The aim of filtering is to estimate x_t given the measurements $y_1, y_2, \cdots, y_{t-1}$, assuming "on-line" processing of the data as it is made available. Smoothing on the other hand, allows state estimation when complete data sets are available and computes the state of system conditional on all observation, past, present and future. Due to the Markovian property of the state equation, smoothing can also be done recursively using the SSM structure. The state estimate $\hat{x}_t$ obtained during the forward pass or KF is updated by the observations $y_{1:n}$. State smoothing is considered to be a backward recursion when the KF is seen as a forward recursion and some quantities computed by the forward pass such as a_t, P_t, K_t, F_t and v_t are stored to be used by the state smoother. Such as combination of a forward and backward pass is also called the Kalman Filter-Smoother (KFS). We can distinguish three types of smoothing problems (Einicke, 2012):

1. Fixed-interval smoothing: Given a fixed interval of observations (possibly complete dataset), the smoothed states are obtained at all times in that interval: $\hat{x}_t = E(x_t|y_{1:n})$ for $t = 1, \cdots, n$.

2. Fixed-point smoothing: states estimates at a fixed point in time are continuously updated using new measurements: $\hat{x}_t = E(x_t|y_s)$, where $s = t+1, \cdots, n$ and t is a fixed positive integer.

3. Fixed-lag smoothing: states x_t are estimated after a fixed numbers of measurements are obtained: $\hat{x}_t = E(x_t|y_{t+s})$, where $t = 1, \cdots, s$ and s is a fixed positive integer.

Depending on the problem studied, the three smoothing techniques offer different "improved" states estimates. In line with the purpose of this thesis, using a complete data set, a fixed-interval smoothing method is preferred. Up to date, a number of interval-fixed smoothing algorithms have been developed such as the Rauch-Tung-Striebel smoother (Rauch et al., 1965), the two-filter Fraser-Potter formula (Fraser and Potter, 1969), etc.

De Jong's cross-validation filter, a fixed-interval type smoothing algorithm, was developed by De Jong (1988). Koopman (1997); Koopman and Durbin (2003) use the cross-validation filter as a smoothing algorithm for their EIKF by modifying it to suit the diffuse initialization. Diffuse smoothing is an added value to the diffuse filtering as it allows to extract initial values if they are needed. Although this thesis is mainly concerned with how to start the filtering process when the initial conditions are unknown, we also present the smoothing recursions to determine the initial values of the system as it is an important contribution to the problem of initial values in general.

For a non-diffuse SSM, the smoothing recursions developed in De Jong (1988) are given by:

Algorithm 3 Smoothing Algorithm

Initialize with $r_n = 0$ and $N_n = 0$

for all $t = n, .., d + 1$ **do**

$L_t = T_t - T_t K_t Z_t$

$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t$

$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t$

$\hat{x_t} = a_t + P_t r_{t-1}$

$\hat{P}_t = P_t - P_t N_{t-1} P_t$

end for

Algorithm 4 Smoothing Algorithm

For $t = n, .., d+1$, apply algorithm 3.

for all $t = d, .., 1$ **do**

Initialize with $r_d^{(0)} = r_d^{(1)} = 0$, $N_d^{(0)} = N_d$ and $N_d^{(1)} = N_d^{(2)} = 0$

if $F_{\infty,t}$ is nonsingular **then**

$$L_{\infty,t} = T_t - T_t K_{\infty,t} Z_t$$
$$r_{t-1}^{(0)} = L_{\infty,t}' r_t^{(0)}$$
$$N_{t-1}^{(0)} = L_{\infty,t}' N_t^{(0)} L_{\infty,t}$$
$$r_{t-1}^{(1)} = Z_t'(F_{\infty,t}^{-1} v_t - K_{*,t}' r_t^{(0)}) + L_{\infty,t}' r_t^{(1)}$$
$$N_{t-1}^{(1)} = Z_t' F_{\infty,t}^{-1} Z_t + L_{\infty,t}' N_t^{(1)} L_{\infty,t} - < L_{\infty,t}' N_t^{(0)} K_{*,t} Z_t >$$
$$F_{\#,t} = K_{*,t}' N_t^{(0)} K_{*,t} - F_{\infty,t}^{-1} F_{*,t} F_{\infty,t}^{-1}$$
$$N_{t-1}^{(2)} = Z_t' F_{\#,t} Z_t + L_{\infty,t}' N_t^{(2)} L_{\infty,t} - < L_{\infty,t}' N_t^{(1)} K_{*,t} Z_t >$$
$$\hat{\alpha}_t = a_t + P_{*,t} r_{t-1}^{(0)} + P_{\infty,t} r_{t-1}^{(1)}$$
$$\hat{P}_t = P_{*,t} - P_{*,t} N_{t-1}^{(0)} P_{*,t} - < P_{\infty,t} N_{t-1}^{(1)} P_{*,t} > - P_{\infty,t} N_{t-1}^{(2)} P_{\infty,t}$$

else

if $F_{\infty,t} = 0$ **then**

$$L_{*,t} = T_t - T_t K_{*,t} Z_t$$
$$r_{t-1}^{(0)} = Z_t' F_{*,t}^{-1} v_t + L_{*,t}' r_t^{(0)}$$
$$N_{t-1}^{(0)} = Z_t' F_{*,t}^{-1} Z_t + L_{*,t}' N_t^{(0)} L_{*,t}$$
$$r_{t-1}^{(1)} = T_t' r_t^{(1)}$$
$$N_{t-1}^{(1)} = T_t' N_t^{(1)} L_{*,t}$$
$$N_{t-1}^{(2)} = T_t' N_t^{(2)} T_t$$
$$\hat{\alpha}_t = a_t + P_{*,t} r_{t-1}^{(0)} + P_{\infty,t} r_{t-1}^{(1)}$$
$$\hat{P}_t = P_{*,t} - P_{*,t} N_{t-1}^{(0)} P_{*,t} - < P_{\infty,t} N_{t-1}^{(1)} P_{*,t} > - P_{\infty,t} N_{t-1}^{(2)} P_{\infty,t}$$

end if

end if

end for

2.3 Extended Kalman Filter

The Kalman filter is contingent on two assumptions: linearity of the models and normality of the distributions. Linearity preserves the Gaussian property of the distributions. When the two assumptions are not met, the distribution of the state estimate is not Gaussian and thus cannot be characterized by its first two moments and thus rendering the KF unusable. This problem is overcome by linearizing the nonlinear evolution and observation models, if they are both nonlinear, around the most current estimate of the state and use the standard KF.

In so far, the discussed Kalman filter and smoother algorithms consider evolution and measurements models updates to be discrete in time. However, many dynamical systems evolve continuously in time and are represented by ordinary differential equations(ODEs).

Two different cases arises:

1. Continuous-Continuous: When both state and measurement equations are characterized by ODEs.
2. Continuous-Discrete or Discrete-Continuous : When one of the equations of the SSM is an ODE or a system of ODEs and the other is discrete equation, with respect to time.

The discrete-discrete context provides a more straightforward setting for developing new algorithms which can later be generalized to the two cases enumerated above in order to depict more realistic real life problems. The next subsection discusses a discrete-discrete extension of the Kalman filter for nonlinear models.

2.3.1 Discrete-Discrete Extended Kalman Filter

A nonlinear SSM can be written as follows,

$$x_{t+1} = T_t(x_t) + R_t\epsilon_t, \tag{2.12a}$$

$$y_t = Z_t(x_t) + \zeta_t. \tag{2.12b}$$

The linearization is done via a first order Taylor expansion of the nonlinear functions $T(\cdot)$ and $Z(\cdot)$ performed around a_t and $a_{t|t}$, the latest state estimates such that:

$$T_t(x_t) \approx T_t(a_{t|t}) + \dot{T}_t \times (x_t - a_{t|t}),$$
$$Z_t(x_t) \approx Z_t(a_t) + \dot{Z}_t \times (x_t - a_t),$$

and

$$\dot{Z}_t = \left.\frac{\partial Z(x)}{\partial x}\right|_{x=a_t}, \qquad \dot{T}_t = \left.\frac{\partial T(x)}{\partial x}\right|_{x=a_{t|t}}. \tag{2.13}$$

Letting,

$$u_t = T_t(a_{t|t}) - \dot{T}_t a_{t|t}, \qquad v_t = Z_t(a_t) - \dot{Z}_t a_t,$$

then Equations (2.12a) and (2.12b) can be reformulated to resemble a linear SSM with inputs :

$$x_{t+1} = T_t x_t + u_t + R_t \epsilon_t, \tag{2.14}$$
$$y_t = Z_t x_t + v_t + \zeta_t.$$

The standard Kalman filter is modified to accommodate the approximation as in Algorithm 5:

Algorithm 5 Extended Kalman filter

Initialize Kalman filter with (a_1, P_1)

for all $t = 1, ..., n$ **do**

$v_t = y_t - Z_t(a_t)$

$F_t = \dot{Z}_t P_t \dot{Z}_t' + H_t$

$K_t = P_t \dot{Z}_t' F_t^{-1}$

$a_{t|t} = a_t + K_t v_t$

$a_{t+1} = T_t(a_{t|t})$

$P_{t|t} = P_t - K_t F_t K_t'$

$P_{t+1} = \dot{T}_t' P_{t|t} \dot{T}_t' + R_t Q_t R_t'$

end for

The extended Kalman filter performs well in general for nonlinear models with Gaussian noises but tends to underestimate the state covariance when the nonlinearities are severe. The extended Kalman filter with diffuse initial conditions is quite similar to the exact initial Kalman filter. The filter is split into a diffuse component and a proper component. The condition for the automatic collapse of the EKF remains the disappearance of the term associated with κ. The algorithm for the diffuse EKF is given by Algorithm 6. Nonlinear smoothing is mostly identical to the standard linear smoother except for the adjustment required for the nonlinear models. The transition and observation matrices of the linear models are replaced by their respective Jacobians.

Algorithm 6 exact initial extended Kalman filter

Initialize the diffuse extended Kalman filter with (a_1, A), (P_∞, P_*)

for all $t = 1...d \leq n$ **do**

if $F_{\infty,t}$ is nonsingular **then**

$v_t = y_t - Z_t(a_t)$

$F_{\infty,t} = \dot{Z}_t P_{\infty,t} \dot{Z}_t'$

$F_{*,t} = \dot{Z}_t P_{*,t} \dot{Z}_t' + H_t$

$K_{\infty,t} = P_{\infty,t} \dot{Z}_t' F_{\infty,t}^{-1}$

$K_{*,t} = (P_{*,t} \dot{Z}_t' - K_{\infty,t} F_{*,t}) F_{\infty,t}^{-1}$

$P_{\infty,t|t} = P_{\infty,t} - K_{\infty,t} F_{\infty,t} K_{\infty,t}'$

$P_{*,t|t} = P_{*,t} - K_{\infty,t} \dot{Z}_t P_{*,t}' - K_{*,t} \dot{Z}_t P_{\infty,t}'$

$P_{\infty,t+1} = \dot{T}_t P_{\infty,t|t} \dot{T}_t'$

$P_{*,t+1} = \dot{T}_t P_{*,t|t} \dot{T}_t' + R_t Q_t R_t'$

$a_{t|t} = a_t + K_{\infty,t} v_t$

$a_{t+1} = T_t(a_{t|t})$

else

if $F_{\infty,t} = 0$ **then**

$v_t = y_t - Z_t(a_t)$

$F_{*,t} = \dot{Z}_t P_{*,t} \dot{Z}_t' + H_t$

$K_{*,t} = P_{*,t} \dot{Z}_t' F_{*,t}^{-1}$

$P_{\infty,t|t} = P_{\infty,t}$

$P_{*,t|t} = P_{*,t} - K_{*,t} \dot{Z}_t P_{*,t}'$

$P_{*,t+1} = \dot{T}_t P_{*,t|t} \dot{T}_t^{'} + R_t Q_t R_t'$

$a_{t|t} = a_t + K_{*,t} v_t$

$a_{t+1} = T_t(a_{t|t})$

end if

end if

end for

for all $t = d + 1...n$ **do**

Use the algorithm 5

end for

2.3.2 Continuous-Discrete Extended Kalman Filter

Let us now consider the continuous-discrete case and reformulate the state space equations such that observations of the systems are sampled at discrete time points t_k with $k = 1, 2, ..., n$ but the state evolves continuously with respect to time. Assumptions about noise mentioned in section 2 apply, that is that evolution and measurements noise are white noise processes and are serially and mutually independent. Although the functions $T(\cdot)$ and $F(\cdot)$ may be linear, it is seldom the case when modeling real-world systems. Thus we will assume in this section, that the state and measurement functions are nonlinear and require the use of an extended Kalman filter.

$$\begin{aligned} \dot{x}(t) &= T(x(t), t) + R(t)\epsilon(t), \quad \epsilon_t \sim N(0, Q_t) \\ y_k &= Z(x(t_k), t_k) + \zeta_k, \quad \zeta_k \sim N(0, H_k), \quad k = 1, ..., n \end{aligned} \tag{2.15}$$

Here $\dot{x}(t)$ is the ODE or system of ODEs that represent the state of the system at any time t. The continuous-discrete EKF (CD-EKF) is similar to its discrete-discrete counterpart in the measurement update steps of the filter. The time update is slightly different due to the time continuity; the state and its covariance matrix are propagated between the previous estimate and the current one using numerical integration schemes to evaluate the system in the interval $t_k < t < t_{k+1}$. The value of the state at $t = t_{k+1}$ is retained as the current state estimate given $y_{1:k}$.

$$\dot{a}(t) = T(a(t), t), \tag{2.16}$$

$$\dot{P}(t) = \dot{T}_t P_t + P_t \dot{T}_t' + R_t Q_t R_t', \tag{2.17}$$

in which $\dot{T}_t = \left.\frac{\partial T(x)}{\partial x}\right|_{x=a_{t|t}}$, is the Jacobian of the evolution model.

Numerically, Equations (2.16) and (2.17) are solved using an ODE solver such as ode45 in MATLAB if the differential equations are not stiff, otherwise solvers like ode15s will result in more stable solutions. At each time update t_{k+1}, the odesolver is given the measurement updated state estimate $a_{t_k|t_{k-1}}$ and covariance estimate $P_{t_k|t_{k-1}}$ as an initial condition and the last value of the solutions of (2.16) and (2.17) on the interval $t_k < t < t_{k+1}$ are ascribed to the state and covariance estimate respectively.

The CD-EKF is given below:

Algorithm 7 Continuous-discrete extended Kalman filter

Initialize Kalman filter with (a_1, P_1)

for all $k = 1, ..., n$ **do**

$$v_k = y_k - Z_k(a_k)$$
$$F_k = \dot{Z}_k' P_{t_k} \dot{Z}_k' + H_k$$
$$K_k = P_{t_k} \dot{Z}_k' F_k^{-1}$$
$$a_{t_k|t_k} = a_{t_k} + K_k v_k$$
$$\dot{a}_{t_{k+1}} = T(a_{t_k|t_k}, t)$$
$$P_{t_k|t_k} = P_{t_k} - K_k F_k K_k'$$
$$\dot{P}_{k+1} = \dot{T}_k P_{t_k|t_k} + P_{t_k|t_k} \dot{T}_k{}' + R_t Q_t R_t'$$

end for

The CD-EKF can be easily transformed to a diffuse CD-EKF by splitting it into two parts like we have done before for the EKF. Likewise for the Kalman smoother although some changes have to be made since the time update is now continuous. For $t = n, \ldots, d+1$, the smoothing recursions for the non diffuse extended smoother are given by :

$$L_t = \Psi_t - \Psi_t K_t Z_t, \quad r_{t-1} = \dot{Z}_t' F_t^{-1} v_t + L_t' r_t, \quad N_{t-1} = \dot{Z}_t' F_t^{-1} \dot{Z}_t + L_t' N_t L_t, \quad (2.18)$$
$$\hat{x}_t = a_t + P_t r_{t-1}, \quad \hat{P}_t = P_t - P_t N_{t-1} P_t,$$

initialized by $r_n = 0$ and $N_n = 0$.

In which, $\Psi = \dot{T}_t$ is the jacobian of the nonlinear evolution function in equation (2.16). Since $T(\cdot)$ is a function that maps $x(t)$ to its derivative $\dot{x}(t)$, so does its Jacobian. Hence, Ψ is integrated by Monte Carlo (MC) integration in the interval $t_k < t < t_{k+1}$. This technique is generally used for multi-dimensional integration problems as it usually provides more accuracy than repeated "dimension-by-dimension" integrations using one-dimensional methods such as trapezoidal or simpson rule. Moreover, as it names suggests, MC integration is based on random numbers and allows to evaluate the integrand at randomly generated points on an interval $[a, b]$. Consider the one dimensional function $f(x)$, then Monte Carlo integration is given by

$$A = \int_a^b f(x)dx = \frac{b-a}{N} \sum_{i=1}^{N} f(x_i) \quad (2.19)$$

where x_i can be sampled from the uniform distribution between a and b. In this particular case, the jacobian $\Psi(t)$ was integrated over the interval $[t_k, t_{k+1}]$ with randomly generated time points $t_i \sim U[t_k, t_{k+1}]$.

After the integration of the jacobian, it can now be used in the smoothing algorithm mentioned above. When $F_{\infty,t}$ is non-singular, the diffuse recursions needed to compute the smoothed state and variance are given by:

$$\begin{aligned}
L_{\infty,t} &= \Psi_t - \Psi_t K_{\infty,t} \dot{Z}_t, \\
r_{t-1}^{(0)} &= L_{\infty,t}^{'} r_t^{(0)}, \\
r_{t-1}^{(1)} &= \dot{Z}_t^{'}(F_{\infty,t}^{-1} v_t - K_{*,t}^{'} r_t^{(0)}) + L_{\infty,t}^{'} r_t^{(1)}, \\
N_{t-1}^{(0)} &= L_{\infty,t}^{'} N_t^{(0)} L_{\infty,t}, \\
N_{t-1}^{(1)} &= \dot{Z}_t^{'} F_{\infty,t}^{-1} \dot{Z}_t + L_{\infty,t}^{'} N_t^{(1)} L_{\infty,t} - < L_{\infty,t}^{'} N_t^{(0)} K_{*,t} \dot{Z}_t >, \\
N_{t-1}^{(2)} &= \dot{Z}_t^{'} F_{\#,t} \dot{Z}_t + L_{\infty,t}^{'} N_t^{(2)} L_{\infty,t} - < L_{\infty,t}^{'} N_t^{(1)} K_{*,t} \dot{Z}_t >,
\end{aligned} \tag{2.20}$$

where $F_{\#,t} = K_{*,t}^{'} N_t^{(0)} K_{*,t} - F_{\infty,t}^{-1} F_{*,t} F_{\infty,t}^{-1}$ and $\langle W \rangle = W + W'$ for a square matrix W.

Otherwise, for a singular $F_{\infty,t}$, we have:

$$\begin{aligned}
L_{*,t} &= \Psi_t - \Psi_t K_{*,t} \dot{Z}_t, \\
r_{t-1}^{(0)} &= \dot{Z}_t^{'} F_{*,t}^{-1} v_t + L_{*,t}^{'} r_t^{(0)}, \\
r_{t-1}^{(1)} &= \Psi_t^{'} r_t^{(1)}, \\
N_{t-1}^{(0)} &= \dot{Z}_t^{'} F_{*,t}^{-1} \dot{Z}_t + L_{*,t}^{'} N_t^{(0)} L_{*,t}, \\
N_{t-1}^{(1)} &= \Psi_t^{'} N_t^{(1)} L_{*,t}, \\
N_{t-1}^{(2)} &= \Psi_t^{'} N_t^{(2)} \Psi_t.
\end{aligned} \tag{2.21}$$

In both cases, the recursions are initialized by $r_d^{(0)} = r_d^{(1)} = 0$, $N_d^{(0)} = N_d$, $N_d^{(1)} = N_d^{(2)} = 0$ and the formulas for the state and its covariance are the same:

$$\begin{aligned}
\hat{\alpha}_t &= a_t + P_{*,t} r_{t-1}^{(0)} + P_{\infty,t} r_{t-1}^{(1)} \\
\hat{P}_t &= P_{*,t} - P_{*,t} N_{t-1}^{(0)} P_{*,t} - < P_{\infty,t} N_{t-1}^{(1)} P_{*,t} > - P_{\infty,t} N_{t-1}^{(2)} P_{\infty,t}
\end{aligned} \tag{2.22}$$

3 Markov Chain Monte Carlo Methods

Bayesian inference is a branch of statistics in which prior belief or knowledge is used in conjunction with data to deduce certain properties of a distribution or a population. The prior information or belief is called the prior distribution $p(\theta)$. When inferring the parameters θ of a given model, they are considered to be random variables. Bayes' rule allows for an update of prior belief about the parameters using the information gathered through the data. The goal of Bayesian parameter estimation is to obtain this updated distribution called the posterior distribution $\pi(\theta|y)$, from which different point estimates such as the maximum a posteriori Estimator (MAP) or the maximum likelihood estimator(MLE) can be obtained. By Bayes' rule, we have that

$$\pi(\theta|y) \propto p(\theta) \times l(y|\theta) \tag{3.1}$$

Here $l(y|\theta)$ represents the likelihood function containing the information provided by the data. Since the posterior distribution must integrate to one, Equation (3.1) becomes:

$$\pi(\theta|y) = \frac{l(y|\theta)p(\theta)}{\int l(y|\theta)p(\theta)d\theta}, \tag{3.2}$$

where the denominator is called the normalization constant. In a high dimension parameter space scenario however, this integral can prove very difficult to compute either analytically or using classical numerical integration methods. Using Monte Carlo methods, we can sample from the distribution of the parameters without directly computing the normalizing constant. The term Markov chain here refers to the fact that the sample chain is constructed in such a way that each realization only depends on the previous one. Markov Chain Monte Carlo (MCMC) algorithms are ergodic and thus allow that, for large enough samples, the sampled distribution will be close enough to the target distribution,i.e., the posterior distribution. We will introduce here two MCMC algorithms: Metropolis Algorithm and Adaptive Metropolis Algorithm.

3.1 Metropolis Algorithm

Developed by Metropolis et al. (1953), the Random Walk Metropolis(RMW) is without doubt one of the most popular MCMC algorithms in use today. The RWM

uses a simple acceptance/rejection rule to progressively converge towards the target distribution. The pseudo-algorithm goes as follows:

Algorithm 8 Metropolis Algorithm

Choose a starting value θ_0 and a sample size N

for all $t = 1, ..., N$ **do**

Choose a *suitable* proposal distribution $q(\hat{\theta}|\theta_{t-1})$

Sample a new candidate $\hat{\theta}$ from chosen proposal

Compute the ratio $r = \frac{\pi(\hat{\theta})}{\pi(\theta)}$

if $\hat{\theta}$ is accepted with acceptance probability r **then**

$\theta_n = \hat{\theta}$

else

if $\hat{\theta}$ is rejected **then**

$\theta_t = \theta_{t-1}$

end if

end if

end for

A key feature of this algorithm, is that we do not have to deal with the normalizing constant which cancels itself when taking the acceptance ratio r in Algorithm 8 . Note that in the RWM, the proposal distribution must be symmetric, *i.e*, the probability of getting $q(\theta_1|\theta_2)$ is the same as that of getting $q(\theta_2|\theta_1)$. A variant of the RWM where the proposal is not symmetric is the *Metropolis-Hastings* algorithm. A lot of attention is payed to the scaling of the proposal distribution as it determines the outcome of the sampling. If the variance is too small, then new candidates will mostly be accepted but in the close vicinity of the previous one and the chain would take long to converge. Otherwise, the acceptance rate is too low and the sampler stays still for a long period of time.

3.2 Adaptive Metropolis Algorithm

As mentioned earlier, the choice of a proper proposal is key in order to obtain reasonable results. The goal of Adaptive Metropolis algorithms is to tune the proposal distribution to match the target distribution both in size and in spatial orientation. Different algorithms have been developed for this purpose like Gilks et al. (1998)

and Brockwell and Kadane (2005) that perform the adaptation at regeneration times but, in this thesis, we will pay a particular attention to the algorithm developed by Haario et al. (2001). For a more effective sampling, the adaptive Metropolis (AM) adjusts the shape and size of the proposal distribution by taking into consideration all the previous states. Note that this renders the chain non Markovian. However, Haario et al. (2001) show that its ergodic property is preserved.

Assuming that we already have $(\theta_0, \theta_1, .., \theta_{t-1})$ samples then, the new candidate will be drawn from a proposal distribution centered at θ_{t-1} and covariance $C_n = s_d Cov(\theta_0, \theta_1, .., \theta_{t-1}) + s_d \epsilon I_d$. Here s_d is the scaling factor and $\epsilon > 0$ ensures that the covariance matrix remains positive definite. No restrictions are imposed on the pre-adaptation period t_0 , however its length reflects our trust in the initial proposal covariance C_0. If the latter has been defined as per some a priori knowledge, the length of the pre-adaptation period might be more lengthy. Otherwise, such a lengthy start might have an effect on the impact of the adaptation on the results (Haario et al., 2001). Thus,

$$C_t = \begin{cases} C_0, & t \leq t_0 \\ s_d Cov(\theta_0, \theta_1, ..., \theta_{t-1}) + s_d \epsilon \mathbf{I}_d, & t > t_0. \end{cases} \tag{3.3}$$

Gelman et al. (1996) established an optimal scaling factor $s_d = 2.38/\sqrt{d}$ for Gaussian targets and Gaussian proposals. A combination of the optimal scaling factor and a Jacobian-based covariance matrix can serve as an initial proposal distribution before the AM starts the tuning.

The update of the covariance can be done with all the previously sampled states $[\theta_0, \theta_1, ..., \theta_t]$ or with an increment such as $[\theta_{t/2}, ..., \theta_t]$. However, as the simulation goes, the AM no longer gathers new information from the previous points of the chain and returns to being a RWM.

3.3 MCMC Convergence Diagnostics

Markov chain is said to have converged when it has reached a stage where it is deemed a representative sample of the underlying stationary distribution. It is often assessed by how well the chain has mixed. By the mixing of the chain we mean the

degree to which the Markov chain explores the support of the posterior distribution. There are a number of visual and statistical tools to assess convergence. Below are some of the graphical tools used in MCMC convergence diagnostics:

- Trace plots: or time series plot is a graph showing the values of a parameter at each iteration of the chain. Convergence is often assessed given the mixing of the chain. Parameter values that move from one region of the parameter space to the other in one step often indicate good mixing. Chains that remain still for long periods of time suggest that the proposal distribution is too large causing too many candidates to be rejected. On the other, wavy looking chains are a sign of slowly moving sampler that will take long to explore the parameter space.

- Two-dimensional parameter plots: are pairwise scatter plots made for every possible pair of parameters. They reveal correlations between parameters that might slow down the mixing of the chain. When high correlations are detected, to improve the mixing of the chain, model simplification or re-parametrization can be considered.

- Autocorrelation function plots: are not of themselves a convergence diagnostic test, but helps assess how far apart are two uncorrelated samples of a chain. The shorter the distance between uncorrelated parameter values the better.

In addition to graphical methods, statistical tests can also be used in convergence assessment. Two such diagnostic test that will be covered in this thesis are the Geweke and the integrated auto-correlation time test. The Geweke test compares the mean of the first 10% percent of the chain's samples to the second half of the chain, that can regarded as having converged. If the difference is not significant then we consider the chain to have converged in the first 10% samples. The integrated auto-correlation time test on the other hand, deals with levels of autocorrelation within the chain. High levels indicate poor mixing when low levels indicate otherwise. It is useful when comparing the efficiency of different samplers.

Let's now consider the example of the Phocine Distemper Virus (PDV) that spread in the seal population of England in 1988. We are going to estimate the parameters of the model using MCMC methods. The dynamics of the disease are represented by SIR model formulated as a system of differential equations with $[S, I, R, H] = [3400, 10, 0, 0]$ as initial conditions.

$$\begin{aligned} \frac{dS}{dt} &= -\alpha IS \\ \frac{dI}{dt} &= \alpha IS - \beta I \\ \frac{dR}{dt} &= (1-f)\beta I \\ \frac{dH}{dt} &= f\beta I, \end{aligned} \tag{3.4}$$

where α is the contact rate, β the removal rate and f the survival rate. All three parameters are estimated using the two MCMC methods introduced. Trace plots of the sampler path are a useful to assess the convergence of a MCMC chain. A well mixing chain has a relatively constant mean and variance and seems to be jumping from one remote region to another rather quickly. The trace plots for both algorithms will also help evaluate the impact of the choice of the proposal distribution. Figure 3 shows a well mixed chain for the first parameter α but very bad ones for the other two.

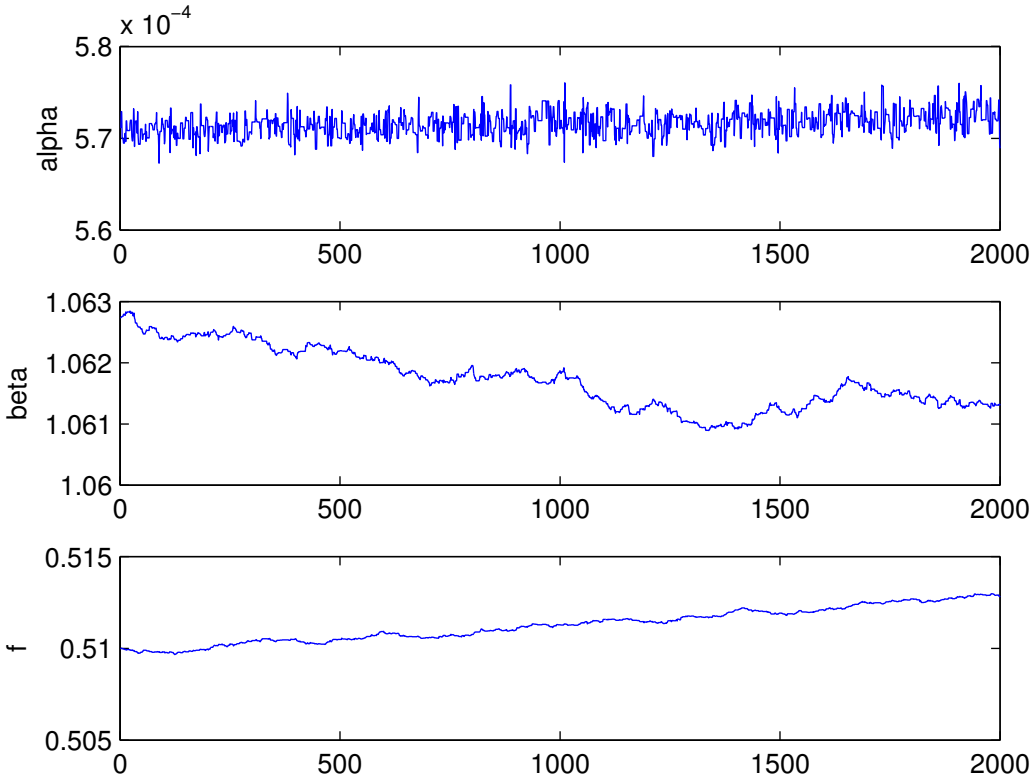

Figure 3: Random Walk Metropolis MCMC Algorithm

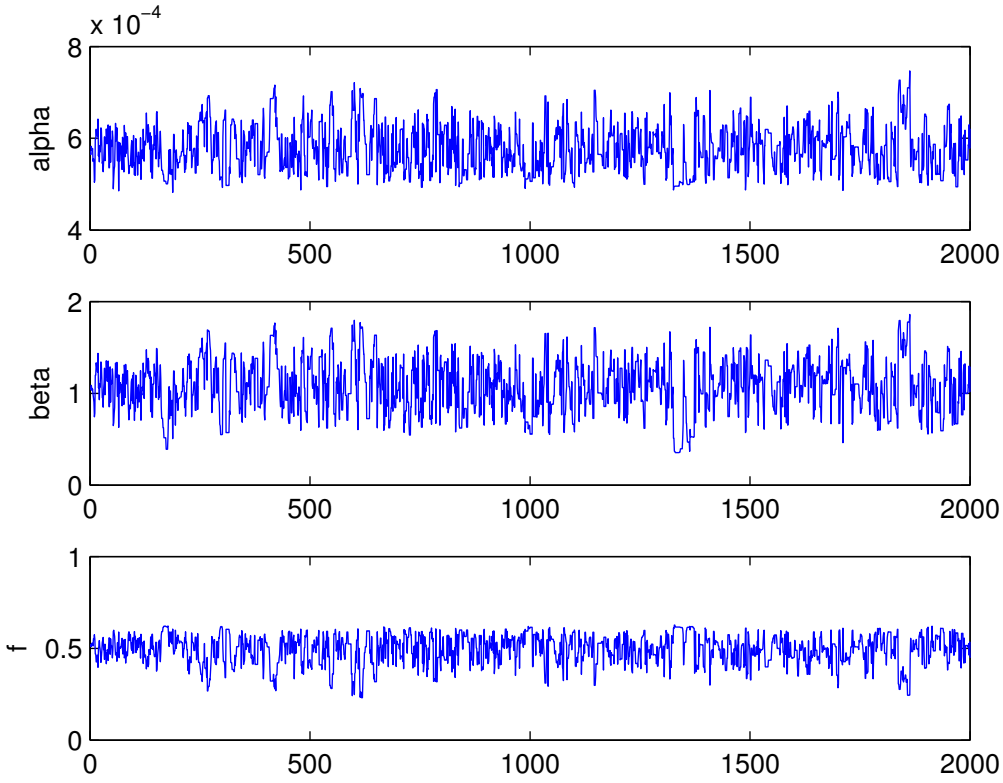

Figure 4: Adaptive MCMC Algorithm

The β and f chain are mixing very slowly and show no sign of convergence. After incorporating a covariance update, we can see a better mix in both chains in Figure 4. Also when looking at Table 3.3, we can see that the autocorrelation time for the AM is much smaller than for the RWM, showing that the AM is more efficient than a simple the RWM.

Methods	Parameters	Mean	Std	tau	Geweke
AM	α	5.86×10^{-4}	3.79×10^{-6}	61.7	0.987
	β	1.12	0.3	69.7	0.942
	f	0.48	0.08	61.2	0.979
RWM	α	5.95×10^{-4}	2.99×10^{-5}	1767.2	0.965
	β	1.18	0.14	1929	0.912
	f	0.47	0.04	1900.6	0.931

Table 1: Comparison of Chain Statistics between RWM and AM Algorithms

The pairwise parameter plots in Figure 5 and 6 both show strong correlations between the parameters but Figure 5 reveal more uncertainty in the parameters.

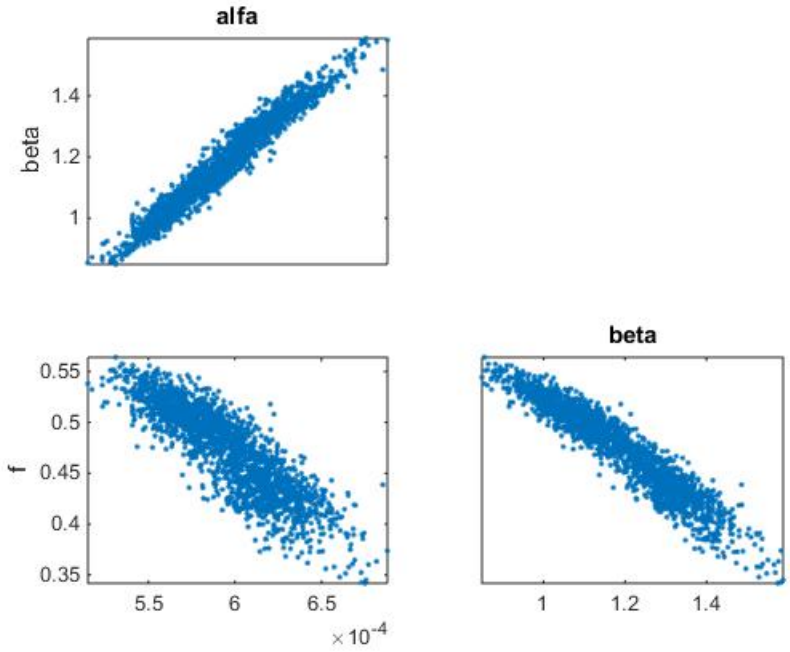

Figure 5: Pairwise scatter plots of RWM chain

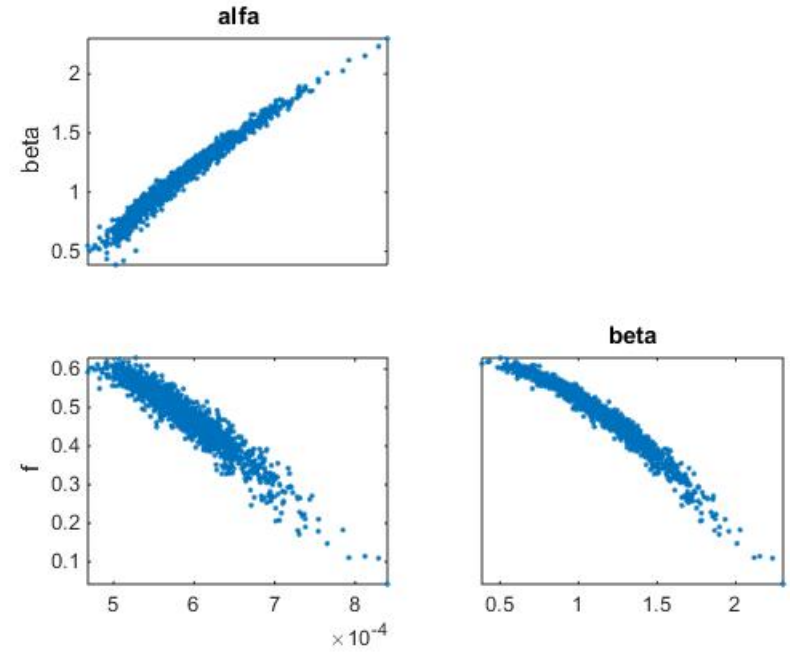

Figure 6: Pairwise scatter plots of AM chain

4 Application: Initial Conditions in Epidemiological modeling

In mathematical epidemiology, deterministic compartmental models are often used to reflect the dynamics of infectious diseases. The simplest and earliest compartmental model, the Susceptible-Infected-Recovered (SIR) model, is attributed to Kermack and McKendrick (1927) and has been used to model many infectious diseases to date such as measles, rubella, influenza, etc. It segments the population under study in compartments that are representative of the major stages of the studied disease. The dynamics from one compartment to another are characterized by differential equations whose parameters, such as recovery rate or death rate cannot be directly quantified and thus have to be estimated.

Parameter estimation involves solving an ordinary differential initial value problem and minimizing a cost function to obtain optimal parameters that fit the observations from the studied dynamical system , in this case, the spread of an infectious disease. More often than not, the dimension of the problem, that is the number of parameters to be estimated, is quite large. Most of the time, analytical solutions to the differential equations cannot be found and must be approximated using diverse numerical schemes. These numerical solutions are sensitive to the inputs of the model, such as initial guesses for the parameters as well as the states and can lead to inaccurate results if badly defined. When there is insufficient knowledge of the initial state of the system, initial values can be considered as additional parameters to be inferred from the data (Bowong and Kurths, 2010), which might prove challenging in highly dimensional models.

The second most discussed method for dealing with uncertainty in the initial conditions is to attribute a distribution to the initial conditions, usually a Beta distribution Beta(α, β) (Kegan and West, 2005; Omar and Hasan, 2012). The different states of an SIR model have probability density functions and are represented by their mean and variance at any time t for certain values of α and β. Although, this method recognizes and accommodates the uncertainty in the initial conditions, it does not provide a way to identify them and requires the introduction of two more parameters, α and β, to take into consideration.

This thesis proposes to approach the initial value problem by accounting for the lack of knowledge through one of the diffuse Kalman-type filters introduced in the previous sections. The filtering and smoothing algorithms will be applied on 1995

Ebola Outbreak in Kikwit, ex-Zaire, now DRC to identify the initial values of the system as well as estimate the parameters. The differential equations of the system, as will be seen in the next section, intuitively require a nonlinear continuous filter.

4.1 1995 Ebola Outbreak in Kikwit, DRC

Disease and model description

Ebola Hemorrhagic Fever (EHF) is a viral disease with a high fatality rate that first simultaneously appeared in 1976, in ex-Zaïre (now DRC) and in Sudan. It was named Ebola after the name of the river in the vicinity of which it first appeared in Zaïre. The disease is caused by the Ebola virus, five strains of which have been identified up to date. Only one of the virus' strains, Ebola-Reston, is non-lethal to humans although they can become infected with it. The rest of the strains, Ebola-Zaire, Ebola-Sudan, Ebola-Ivory Coast and Ebola-Bundibugyo, are often fatal to humans.

Some of the symptoms of EHF are: sudden onset of fever, intense weakness, vomiting, acute headache and diarrhea. Early on, Ebola symptoms resemble a certain number of other Hemorrhagic fevers which makes its detection difficult. The disease is known to be transmitted to people from gorillas and chimpanzees, although in some regions without apes, humans can also be infected by fruit bats. It propagates in the population via contact with bodily fluids such as blood, semen, saliva, etc., of an infectious person. Once infected, the incubation period is formally between 1 to 21 days, although on average, symptoms are known to appear in the first 10 days after the infection. When the symptoms have appeared, the individual is now infectious, that is, can spread the disease. Up to date, there is no known cure or vaccine for Ebola, thus once infected a person can either die or recover.

Ebola can be roughly captured by SEIR, where E stands for the latency or exposed period when infected individuals do not yet show any sign of the disease. In a population of N individuals, the susceptible S are the proportion of the population at risk, the exposed E have already contracted the disease, the infectious I are the symptomatic and the removed R are those who either survived the disease or the casualties.

This type of SEIR model is quite general and is subject to change according to the different constraints such as control measures, human migration, etc. The Ebola model used in this thesis was developed by Ndaguza et al. (2013) based on the

earlier model by Chowell et al. (2004) but modified to take into account the two data sets available (onset and death data) for the 1995 Kikwit outbreak in DRC. The model is based on the following compartmental model:

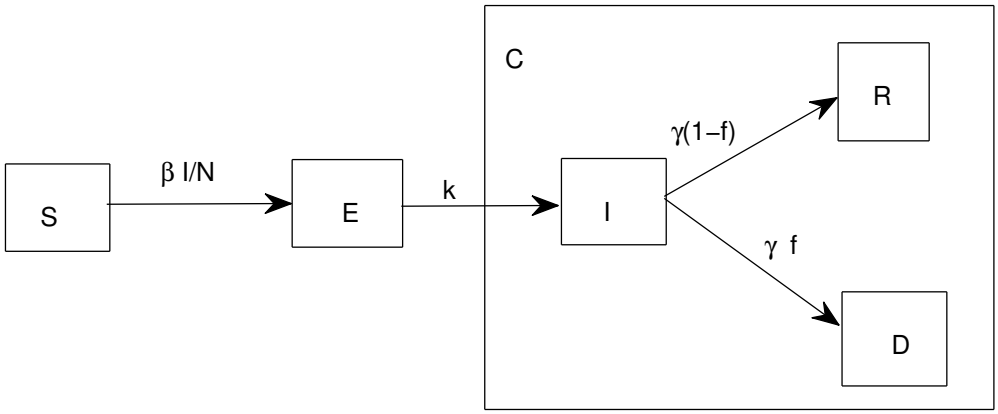

Figure 7: Ebola Dynamics Flowchart

The parameters of the model are transition rates describing the evolution of the disease within the population. β, the transmission rate, is the rate at which susceptibles have adequate contact with infectious individuals and move to the compartment E. Likewise, after a number of days κ, the exposed become infectious and progress to I at the rate $1/\kappa$. From the compartment I, an individual can either recover or die. If we consider γ to be the rate at which infectious individuals leave I for compartment D at the death rate f or for compartment R at the recovery rate $1-f$, we obtain the flowchart above. Note that it is extremely rare that individuals who have contracted Ebola and recovered from it become infected again during the same epidemic, thus the disease cycle is closed.

The Ebola flowchart in Figure 7 can be translated to a system of deterministic

ODEs:

$$\begin{aligned}
\dot{S} &= -\beta(t)S\frac{I}{N}, \\
\dot{E} &= \beta(t)S\frac{I}{N} - \kappa E, \\
\dot{I} &= \kappa E - \gamma I, \\
\dot{R} &= \gamma(1-f)I, \\
\dot{D} &= \gamma f I.
\end{aligned} \tag{4.1}$$

Although, $C(t)$ is not a compartment because it does not represent any stage of the disease as such, it is included in the estimation model to monitor the cumulative number of onset cases. Despite being highly contagious, the implementation of control measures such as quarantine and the use of proper medical equipment can hinder the propagation of Ebola such that the transmission rate β is no longer a constant parameter but becomes a function of time $\beta(t)$ and behaves according to the function below.

$$\beta(t) = \beta_0 + \frac{\beta_1 - \beta_0}{1 + e^{-q(t-\tau)}}, \tag{4.2}$$

In Equation (4.2), the parameter τ marks the time the intervention effect is at 50% and $q > 0$ indicates the rate at which β_0 decay to β_1.

Data

On January 13th 1995, a 42 year old male charcoal worker died at the Kikwit General Hospital after showing symptoms of hemorrhagic disease for approximately a week. He infected, directly and indirectly, several members of his family, among which 13 died in the following 2 months Khan et al. (1995). The virus propagated into the community through the family members of the index case and eventually led to a small nosocomial (i.e, originating in hospital) outbreak at the Kikwit II Maternity in March, 1995. From there it spread to the Kikwit General Hospital and was treated as dysentery for a long time until samples were send to the Center for Disease Control (CDC) where it was definitively diagnosed as Ebola (Khan et al., 1995).

Within 10 days of the diagnostic, preventive measures to control the disease were introduced. Quarantine and isolation wards were set up within the hospital to prevent further spread of the disease. Furthermore, medical teams from international health organizations such as WHO, the CDC, etc arrived in Kikwit and appropriate

protective equipment was given to the medical team on the ground but none to the family members of the quarantined patients. Family members were placed into the isolation ward and moved to the quarantine zone if onset symptoms were noticed.

Although the disease lasted more than 200 days, counting from the first suspected index case, the data was only collected for the time period between March 1st,1995 and July 21st,1995. Of the 316 suspected cases, Khan et al. (1995) only report 291 onset cases and 236 death cases during that time. Overall, the medical corps was the most affected by the disease, with 11% of the fatality cases, of those 7% were nurses.

4.2 Initialization of the diffuse CD-EKF

The system (4.1) was initialized by the diffuse initialization method described in section 2.1.1 and were estimated by continuous-discrete extended Kalman filter-smoother (KFS). For estimation purposes, the compartment $R(t)$ representing the recovering individuals, is replaced by the compartment $C(t)$ since $R(t)$ can be obtained from $D(t)$. The system of differential equations in 4.1 can be formulated as a state-space model as follows:

$$\dot{x}(t) = \begin{pmatrix} \dot{S}(t) \\ \dot{E}(t) \\ \dot{I}(t) \\ \dot{D}(t) \\ \dot{C}(t) \end{pmatrix} = \underbrace{\begin{pmatrix} -\beta(t)S(t)\frac{I(t)}{N} \\ \beta(t)S(t)\frac{I(t)}{N} - kE(t) \\ kE(t) - \gamma I(t) \\ \gamma f I(t) \\ kE(t) \end{pmatrix}}_{T(x(t),t)} + \epsilon_t, \quad \epsilon_t \sim N(0, I) \tag{4.3}$$

$$y_k = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{Z_k} x_k + \zeta_k, \quad \zeta_t \sim N(0, H)$$

The initial state can be divided into two parts to account for diffuse initialization as follows:

$$x_1 = a_1 + A\delta + R_0\epsilon_0, \quad \delta \sim N(0, \kappa I_q), \quad \epsilon_0 \sim N(0, Q_0)$$

Assuming the total population of the Bandundu province to be our only prior knowledge of the system at t_0, the size of the diffuse vector is $q \times 1$, with $q = 4$. The

matrices A and Q_0 are respectively $m \times q$ and $m \times (m-q)$ matrices that consists of columns of the identity matrix I_m where m is the number of states in the system. They are constructed in such a way that all non-zero diagonal elements of $P_{\infty,1} = AA'$ represent the states whose initial conditions we wish to estimate, which are correspondingly made equal to zero in a_1. Hence,

$$a_1 = \begin{pmatrix} S_0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, R_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{4.4}$$

where $S_0 = 5,365,500$. $P_{\infty,1}$ and $P_{*,1}$ are given by:

$$P_{\infty,1} = AA' = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, P_{*,1} = H_0 H_0' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.5}$$

Finally, the parameters used in the model (4.3) are given in Table 2.

Figure 8 shows the fit of the EIKF used to estimate the states. We can see that the diffuse CD-EKF closely follows the data after the slight bump in the beginning at time $t = 3$ when the diffuse CD-EKF collapses and the usual Kalman filter begins.

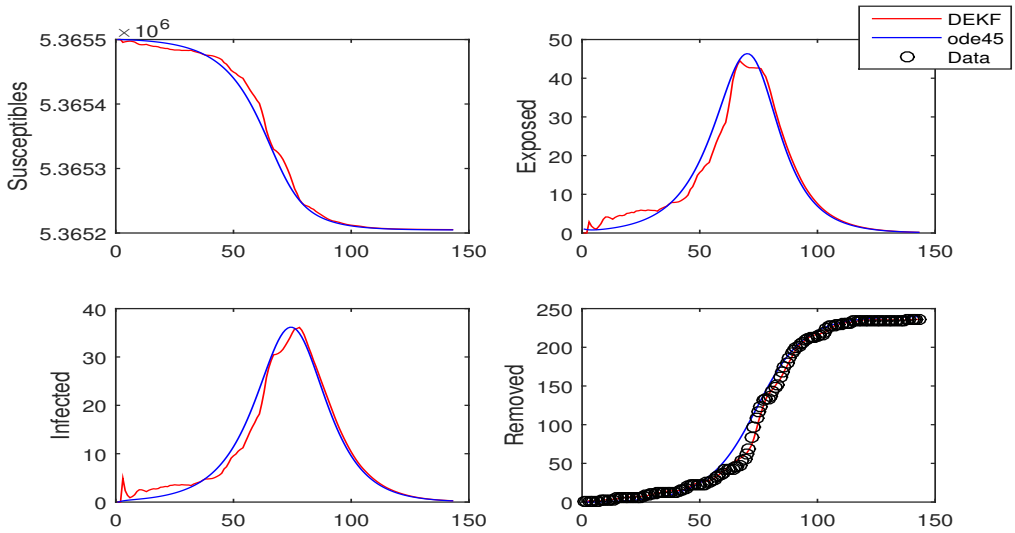

Figure 8: Diffuse CD-EKF(red line) with the numerical solution using Matlab function ode45 (blue line). Black circles are the data corresponding to the Removed(dead) compartment.

4.3 Parameter estimation using MCMC

Parameter estimation is achieved by finding parameters values that maximize the diffuse log-likelihood function in Equation (2.11). Samples from the posterior distribution of the parameters of the Ebola model are obtained via MCMC. Least square estimates given in Table 2 are taken as the starting point of the MCMC algorithm and the sampler is tuned using the Jacobian of the system.

Parameters	Definition	Initial values	Estimates
β_0	contact rate before intervention $(\text{days})^{-1}$	0.3940	0.3914
β_1	contact rate after intervention $(\text{days})^{-1}$	0.0550	0.0565
q	Rate from β_0 to β_1 $(\text{days})^{-1}$	0.1650	0.1660
1/k	Incubation period(days)	1.80	4.9998
γ	Removal rate $(\text{days})^{-1}$	5.0	7.8917
f	Probability of death	0.81	0.7970
τ	Starting time of the intervention (days)	70.0000	70.8796

Table 2: Least square estimates using onset and death data

MCMC statistics for the sampled posterior distributions of parameters are given

in Table 3. According to the Geweke diagnostic test, all parameters seem to have

Parameters	Units	Posterior Mean	Posterior STD	Tau	Geweke
β_0	$(days)^{-1}$	0.34116	0.06965	271.18	0.95076
β_1	$(days)^{-1}$	0.07956	0.04746	386.98	0.81872
q	$(days)^{-1}$	0.38093	0.21291	3562	0.80913
1/k	(days)	1.7029	0.21069	652.89	0.95987
γ	$(days)^{-1}$	4.9879	1.1756	314.25	095309
f	%	83.374	0.08380	1128.6	0.88629
τ	(days)	73.092	1.1529	45.204	0.99971

Table 3: MCMC statistics using diffuse likelihood

converged in the first 10% of the chain. Assessing the convergence graphically, however, reveals that some parameters have identified better than others.

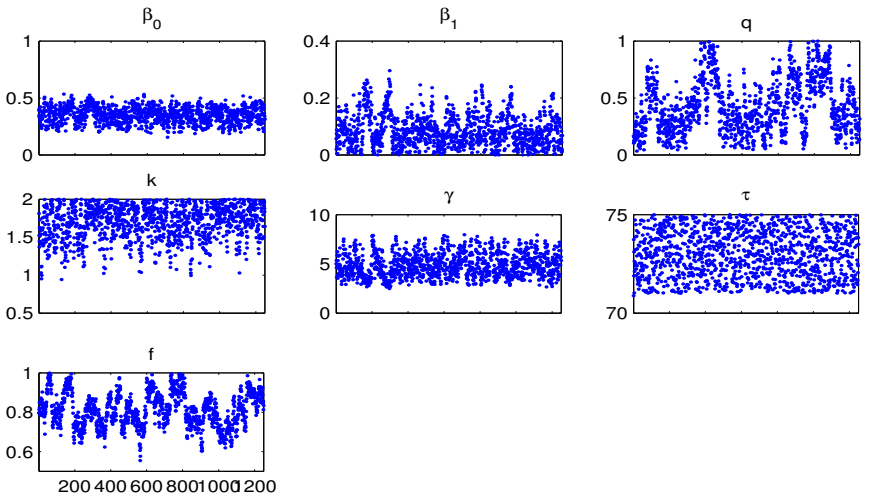

Figure 9: 1D plot of all parameters with an interval of 80 samples with diffuse likelihood

Figure 9 shows a 1D or a trace plot of all parameters. Despite the small number of simulations (approximately 100,000 samples), it seems the parameters β_0, β_1, k and γ have mixed well. We can further see from the 2D plots in Figure 4.3 that there is a clear correlation between the contact rates before and after the intervention (β_0 and β_1) and γ the removal rate parameter. All the other parameters show no strong sign of correlation. The pairwise plot also reveal that the parameter τ does not identify well.

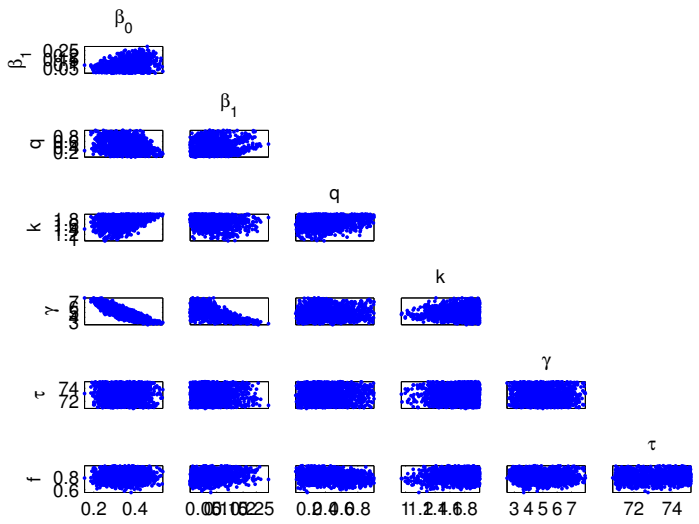

Figure 10: 2D or pairwise plots of all parameters with an interval of 80 samples with diffuse likelihood

Now Table 4 shows the results when running the same MCMC algorithm with an ordinary likelihood.

Parameters	Units	Posterior Mean	Posterior STD	Tau	Geweke
β_0	$(days)^{-1}$	0.34086	0.06422	313.38	0.96214
β_1	$(days)^{-1}$	0.07547	0.04714	512.47	0.76491
q	$(days)^{-1}$	0.47197	0.24147	2934.9	0.68075
1/k	(days)	1.7289	0.19741	470.07	0.93944
γ	$(days)^{-1}$	5.1301	1.1185	493.93	0.87918
f	%	81.278	0.07301	1546.2	0.9379
τ	(days)	73.093	1.1383	39.775	0.99865

Table 4: MCMC statistics using ordinary likelihood

The statistics for the posterior mean and standard deviation are quite similar to those obtained using the diffuse likelihood function, however the integrated autocorrelation time is smaller for most parameters in the diffuse likelihood case than in the ordinary likelihood case. Although, MCMC algorithms with diffuse likelihood provide more or less similar results to those with ordinary likelihood, they are computationally heavier.

Graphical results also indicate that there is not significant difference between the

two samplers as seen in Figures 11 and 12.

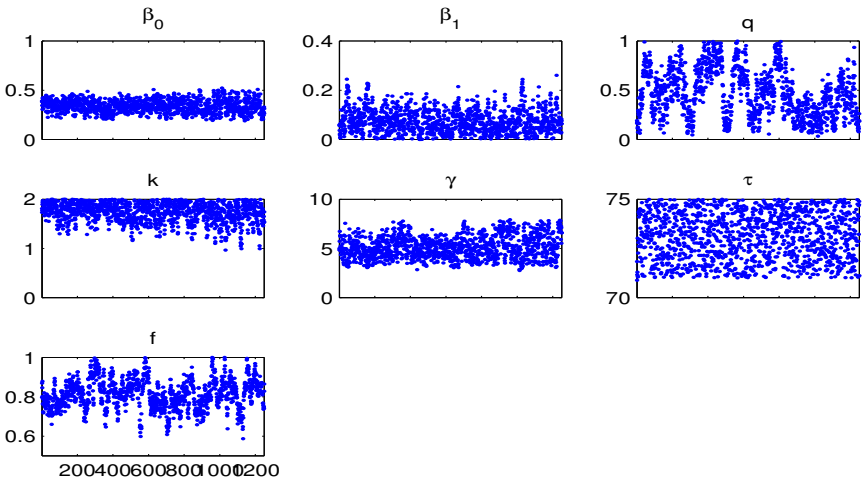

Figure 11: 1D plot or traceplot of all parameters with an interval of 80 samples with ordinary likelihood

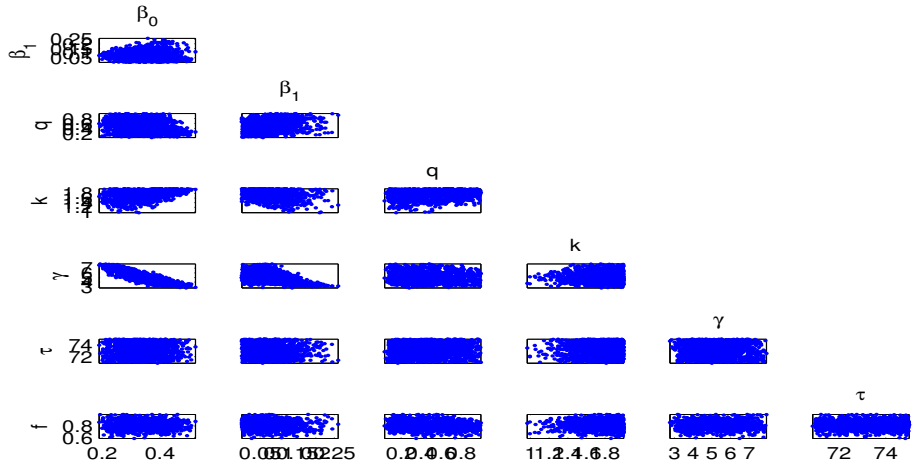

Figure 12: 2D or pairwise plots of all parameters with an interval of 80 samples with ordinary likelihood

5 Conclusion

The Kalman filter and its various extensions all assume that the distribution of the initial state is known, however this is seldom the case in reality. Diffuse initialization provides a way to start the filter without guessing or directly attributing a large covariance to the initial conditions. The effects of the diffuse initialization often dissipate after a few iterations, d, causing the diffuse Kalman filter to collapse to its standard version. In addition, a diffuse smoothing algorithm allows the retrieval of the initial values when they are of interest in their own right to the modeler.

Extensions of the diffuse initialization procedure to continuous nonlinear models are for the most part straightforward and easy to implement in the forward pass or filtering phase. The smoothing recursions, however, involve the backward integration of the jacobian of the transition matrix which can be done using Monte Carlo integration.

The second objective of this thesis, which was to study the feasibility of parameter estimation when dealing with unknown or partially unknown initial conditions was achieved using the diffuse log likelihood. The main example taken in this thesis was the 1995 Ebola Outbreak in Kiwkit, former Zaire, now DRC. The parameters of the epidemiological model are estimated using an adaptive MCMC algorithm with a diffuse likelihood function. This method provides a relatively good approximation of the posterior distribution of the parameters considering the small number of simulations ran. The MCMC sampler with a diffuse likelihood seems, however, computationally heavier than its ordinary counterpart. On the other hand, considering that results are a quite similar, this trade off between computational time and the ability to start parameter estimation without prior knowledge of the state of the dynamical system is at the discretion of the researcher, but for low dimensional models or models that do not require long simulations to attain convergence, using MCMC methods coupled with a diffuse likelihood clearly takes the upper hand.

The work done in this thesis can be extended to develop the diffuse initialization techniques for other Kalman-types algorithms such as the unscented Kalman filter or the particle filter.

References

B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, N.J., 1979.

Craig F. Ansley and Robert Kohn. Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *The Annals of Statistics*, 13(4):1286–1316, 1985.

Craig F. Ansley and Robert Kohn. Filtering and smoothing algorithms for state space models. *Computers and Mathematics with Applications*, 18(6-7):515–528, 1989.

S. Bowong and J. Kurths. Modelling tuberculosis and hepatitis b co-infections. mathematical modelling of natural phenomena. *Mathematical Modelling of Natural Phenomena*, 5:196–242, 2010.

A.E. Brockwell and J.B. Kadane. Identification of regeneration times in mcmc simulation, with application to adaptive schemes. *Journal of Computational and Graphical Statistics*, 14:436–458, 2005.

G. Chowell et al. The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda. *Journal of Theoretical Biology*, (229):119â“126, 2004.

S. Chu-Chun-Lin. *Statistical Analysis with the State Space Model*. PhD thesis, University of British Columbia, 1991.

P. De Jong. A cross-validation smoother filter for time series models. *Biometrika*, 75(3):594–600, 1988.

P. De Jong. The diffuse kalman filter. *The Annals of Statistics*, 19(2):1073–1083, 1991.

J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, 2001.

G. Einicke. Continuous-time smoothing, smoothing, filtering and prediction - estimating the past, present and future. 2012. ISBN 978-953-307-752-9.

D. C. Fraser and J. E. Potter. The optimum linear smoother as a combination of two optimum linear filters. *IEEE Transactions on Automatic Control*, AC-14(4): 387 â“ 390, 1969.

A. G. Gelman et al. Efficient metropolis jumping rules. *Bayesian Statistics V*, pages 599–608, 1996.

W. R. Gilks et al. Adaptive markov chain monte carlo through regeneration. *Journal of the American Statistical Association*, 93(444):1045–1054, 1998.

H. Haario et al. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

J.D. Hamilton. State-space models. In R. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4. Elsevier Science B.V., Amsterdam, The Netherlands, 1994a.

A.C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, UK, 1989.

A.C. Harvey and G.D.A Phillips. Maximum likelihood estimation of regression models with autoregressive- moving average disturbances. *Biometrika*, 66(1):49–58, 1979.

R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASMEâ“Journal of Basic Engineering*, 82:35–45, 1960.

B. Kegan and W.R. West. Modeling the simple epidemic with deterministic differential equations and random initial conditions. *Mathematical Biosciences*, (194): 217â“231, 2005.

W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics, part i. *Proceedings of the Royal Society of Edinburgh*, (115):700–721, 1927.

A.S. Khan et al. The reemergence of ebola hemorrhagic fever, democratic republic of the congo. *Journal of Infectious Diseases*, (179):76â“86, 1995.

Robert Kohn and Craig F. Ansley. Estimation, prediction, and interpolation for arima models with missing data. Technical Report 24, Statistics Research center, University of Chicago, 1984a.

Robert Kohn and Craig F. Ansley. Estimation, prediction, and interpolation for arima models with missing data. *Journal of the American Statistical Association*, 81(395):751–761, 1986.

S.J. Koopman. Exact initial kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association*, 92(440):1630–1638, 1997.

S.J. Koopman and J. Durbin. Filtering and smoothing of state vector for diffuse state space models. *Journal of Time Series Analysis*, 24(1):85–98, 2003.

S.P. Maybeck. *Stochastic Models, Estimation and Control.* Academic Press, Inc, London, 1979.

N. Metropolis et al. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

D. Ndaguza et al. Statistical data analysis of the 1995 ebola outbreak in the democratic republic of congo. *Afrika Matematica*, 24(1):55–68, 2013.

A.H.A. Omar and A.H. Hasan. Numerical simulations of an sir epidemic model with random initial states. 2012.

H. E. Rauch et al. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445 â“ 1450, 1965.

S. Särkkä. *Bayesian Filtering and Smoothing.* Cambridge University Press, 2013.

F.C. Schweppe. *Uncertain Dynamic Systems.* Prentice-Hall, Englewood Cliffs, N.J., 1973.

List of Tables

List of Figures