Applied Mathematics and Computation 316 (2018) 155–166

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Nonlinear Kalman filtering for censored observations

Joseph Arthur[a], Adam Attarian[b], Franz Hamilton[c,*], Hien Tran[c]

[a] Department of Statistics, Stanford University, Stanford, CA 94305 USA
[b] MIT Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02420 USA
[c] Department of Mathematics and Center for Quantitative Sciences in Biomedicine, North Carolina State University, Raleigh, NC 27695 USA

ARTICLE INFO

Keywords:
Extended Kalman filter
Censored observation
Parameter estimation
Hepatitis C virus (HCV)
Human immunodeficiency virus (HIV)

ABSTRACT

The use of Kalman filtering, as well as its nonlinear extensions, for the estimation of system variables and parameters has played a pivotal role in many fields of scientific inquiry where observations of the system are restricted to a subset of variables. However in the case of censored observations, where measurements of the system beyond a certain detection point are impossible, the estimation problem is complicated. Without appropriate consideration, censored observations can lead to inaccurate estimates. Motivated by previous work on censored filtering in linear systems, we develop a modified version of the extended Kalman filter to handle the case of censored observations in nonlinear systems. We validate this methodology in a simple oscillator system first, showing its ability to accurately reconstruct state variables and track system parameters when observations are censored. Finally, we utilize the nonlinear censored filter to analyze censored datasets from patients with hepatitis C and human immunodeficiency virus.

1. Introduction

The use of data assimilation for the estimation of unobserved model variables and parameters has become standard practice in modern scientific analysis. Kalman filtering [1] and its nonlinear extensions such as the ensemble Kalman filter and extended Kalman filter have gained increasing popularity in application to a variety of problems arising from the physical and biological sciences [2–15].

Of recent interest in the field of biomedicine has been the use of ordinary differential equations to model viral infection dynamics such as human immunodeficiency virus (HIV) and hepatitis C virus (HCV) [16–18]. Such models can provide insights into disease behavior, treatment, and ultimately improve patient outcomes. Their use for the development of patient specific treatment regimens remains an exciting possibility. However, these models are parameterized by a number of unknown parameters and observation of the system is limited to a noisy subset of the dynamical variables.

Several methodologies have developed to handle this problem of state and parameter estimation from noisy observations. In particular, the use of Kalman filtering for joint state and parameter estimation has been the topic of several recent papers [19–22]. Unfortunately, this estimation process is further complicated when we consider that the assays used in viral studies for data collection often have a detection limit beyond which accurate measurements are impossible. We refer to these data as *censored*. Measurements within the detection limit are considered *uncensored*. Ignoring the censored data can lead to bias in the estimates [23]. As such, a proper framework for handling censored observations is required.

* Corresponding author.
E-mail address: fwhamilt@ncsu.edu (F. Hamilton).

http://dx.doi.org/10.1016/j.amc.2017.08.002

Kalman filtering for censored observations has been the topic of several recent works [23–25]. Of particular interest is the method proposed in [24]. There the authors derived an auxiliary set of equations for the Kalman filter which provided a modified Kalman gain and covariance update formula to allow for correct inference given censored measurements. The underlying assumption though was that the system of interest is linear. Unfortunately the majority of physical systems and the models representing them are nonlinear, such as those describing the dynamics of HCV and HIV. Our goal in this article is to extend the methodology presented in [24] to the case of nonlinear system dynamics. We derive a modified version of the extended Kalman filter allowing for the accurate joint estimation of state variables and parameters in nonlinear systems in the presence of censored data.

We validate our proposed nonlinear censored filter first in a synthetic oscillator system where a detection limit for system observation is imposed. We show the fidelity of filter's state variable and parameter reconstruction even when we have partial observability of the system and several of the data are censored. Additionally, we demonstrate the capability of the filter to track system nonstationarity in the form of a drifting parameter whose dynamics are unknown. Motivated by our success in this synthetic example, we consider the difficult problem of state and parameter estimation for clinical viral data. In particular we examine two datasets from an HCV and HIV clinical study, both of which contain numerous censored data in their respective viral load measurements.

In analyzing these clinical datasets, we follow very closely the work done in [26] and [27] for the HCV and HIV data respectively. There, the authors provided a detailed model identifiability analysis for these datasets and performed estimation using the expectation maximization algorithm [28]. Our belief is that the filter should not provide more reliable or accurate estimates than those calculated by expectation maximization, in fact they should be comparable. Therefore we treat the results of [26,27] as "ground truth" and aim to show that the proposed nonlinear censored filter is able to reproduce similar estimates. The true utility of the filter is that it provides sequential estimation allowing for the online joint estimation of state variables and parameters and the possibility of tracking parameters whose values drift over time, both of which expectation maximization are unable to do. These capabilities are of particular interest in the field of personalized medicine where researchers may be analyzing clinical data whose measurements span over several months or years and an accurate and timely estimate of the current system state is necessary for appropriate treatment or intervention.

2. Nonlinear Kalman filtering with censored observations

We assume the following nonlinear system with continuous-time state dynamics and discrete observations

$$\begin{aligned}\dot{x}(t) &= f(t,x) + w(t)\\ z(t_k) &= h(x(t_k)) + v_k,\end{aligned}$$

where x is an n dimensional state vector and z is an m dimensional observation vector. w and v are Gaussian noise terms that correspond to the system and observation noise, with covariances Q and R respectively. Selection of these noise matrices is key to the success of any filtering methodology. The estimation of Q and R, a process known as adaptive filtering, is an active area of research (see [10] and the references within). Since our goal in this manuscript is to examine the problem of state and parameter estimation in nonlinear systems with censored data, we simplify matters by performing offline tuning of these error covariance matrices to obtain optimal filter performance.

Due to the system nonlinearity, the standard Kalman filter can not be applied directly. Several nonlinear filters have developed, such as the ensemble Kalman filter (EnKF) and extended Kalman filter (EKF) [29,30]. Here we consider the EKF, which performs a linearization of the system dynamics at each step of the filter. For a detailed derivation of the algorithm see [31].

The EKF is a sequential estimator that consists of a prediction and update step. We solve the following system

$$\begin{aligned}\dot{\hat{x}} &= f(t,\hat{x})\\ \dot{P} &= PF^T + FP + Q,\end{aligned}$$

with initial conditions $\hat{x}_{k-1}$ and P_{k-1} from t_{k-1} to t_k to compute $\hat{x}_k^-$ and P_k^-, our prior state and covariance matrix estimates. F is the linearization of the system dynamics, namely $F = \nabla f(\hat{x})$. We form the linearization of the observation operator, $H_k = \nabla h(\hat{x}_k^-)$, and then implement the standard Kalman update equations to correct our state and covariance estimates

$$\begin{aligned}\hat{x}_k &= \hat{x}_k^- + K_k\left[z_k - h(\hat{x}_k^-)\right]\\ P_k &= [I - K_k H_k]P_k^-\\ K_k &= P_k^- H_k^T\left[H_k P_k^- H_k^T + R\right]^{-1}.\end{aligned}$$

2.1. Filtering with censored data

In the case of censored data, where the true value of the observation beyond a certain lower or upper detection limit is unknown, the estimation problem is complicated. Treating these censored observations as uncensored measurements leads to inaccurate estimates. In [24], Gabardós and Zufiria addressed this problem of state estimation in the presence of censored