Computational Statistics and Data Analysis 75 (2014) 1–14

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Kalman filter variants in the closed skew normal setting

Javad Rezaie *, Jo Eidsvik

Department of Mathematical Sciences, NTNU, Norway

ARTICLE INFO

Article history:
Received 6 April 2013
Received in revised form 6 January 2014
Accepted 22 January 2014
Available online 30 January 2014

Keywords:
Recursive Bayesian estimation
Closed skew normal distribution
Ensemble Kalman filter
Nonlinear system
Petroleum reservoir

ABSTRACT

The filtering problem (or the dynamic data assimilation problem) is studied for linear and nonlinear systems with continuous state space and over discrete time steps. Filtering approaches based on the conjugate closed skewed normal probability density function are presented. This distribution allows additional flexibility over the usual Gaussian approximations. With linear dynamic systems the filtering problem can be solved in analytical form using expressions for the closed skew normal distribution. With nonlinear dynamic systems an ensemble-based version is proposed for fitting a closed skew normal distribution at each updating step. Numerical examples discuss various special cases of the methods.

1. Introduction

In this paper we consider the filtering problem under non-Gaussian and nonlinear modeling assumptions. The challenge is to characterize the probability distribution of state variables over time given the available information. The underlying model comes from a physical system and is represented as a set of differential or difference equations known as the process model. At each (discrete) time step sensors measure the state variables directly or indirectly. We are interested in assimilating these data with the knowledge imposed by the process model and all previous measurements.

The history of the filtering problem goes back more than 50 years when Kalman proposed his famous filtering solution for linear dynamical systems in optimal control literature (Kalman, 1960). The Kalman filter (KF) has shown extremely useful but has strict assumptions about linearity and Gaussian noise. The Extended Kalman filter (EKF) is an extension handling nonlinearities (Jazwinsky, 1970) but if the nonlinearity of the system is high, this first order approximation diverges. Second order EKF variants have also been proposed but they may have similar challenges. Moreover, we need to calculate the Jacobian and Hessian of the system equations which may not be feasible. For instance, these derivative expressions are rarely available from implicitly formulated process models or black box models.

Particle filters (PF) were proposed to handle general distributions by Monte Carlo sampling. These approaches can approximate any distribution when the number of particles go to infinity (Doucet et al., 2001). In practical systems there are limits to the available computation time, and consequently the number of particles/samples which can be used. Albeit popular in many applications, particle filters may suffer from sample degeneracy when the system dimension increases.

The ensemble Kalman filter (EnKF) was introduced as a sampling representation for very high dimensional systems, see e.g. Evensen (2003), Sakov and Oke (2008) and Evensen (2009). It incorporates the nonlinear process model, whereas a Gaussian approximation is used for the updating with respect to new measurements. This approach has been very useful for practical applications, but the filter solution may be biased or can underestimate uncertainty. Gaussian mixture filters have been suggested to get benefits from both PF and EnKF, see e.g. Stordal et al. (2010) and Rezaie and Eidsvik (2012).

* Corresponding author.
E-mail addresses: rezaie@math.ntnu.no (J. Rezaie), joeid@math.ntnu.no (J. Eidsvik).

http://dx.doi.org/10.1016/j.csda.2014.01.014
0167-9473/

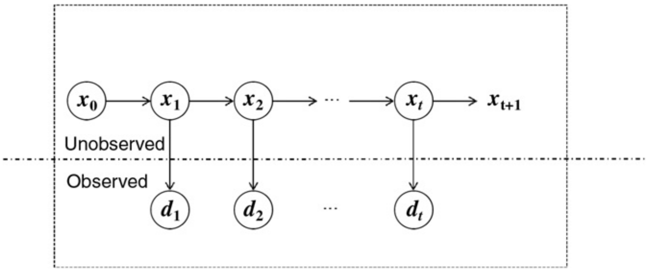

Fig. 1. Graphical representation of state variables $\boldsymbol{x}_t$, at discrete time points $t = 0, 1, \dots$ and observations $\boldsymbol{d}_t$, $t = 1, 2, \dots$. The process model is assumed to follow a Markov structure. The data are assumed to be conditionally independent, given the state variable at the indicated time steps. The filtering problem characterizes the distribution of states over time, given all currently available data.

In this paper we introduce a new filter which captures skewness in the filtering solution. It is easy to implement and some of the mentioned filters are special cases of the suggested approaches. The filter is based on the closed skew normal (CSN) distribution which allows analytical solutions under certain modeling assumptions. This family of filters is named the closed skew normal Kalman filter (CSNKF).

The skew normal (SN) and CSN distribution are extensions of the normal distribution, see e.g. Azzalini and Dalla-Valle (1996) and Gupta et al. (2004). Skewness is incorporated by adding new parameters to the traditional Gaussian formulation (Genton, 2004). The CSN distribution has some useful properties similar to those of the Gaussian distribution such as conjugacy under linear conditioning. Thus, we can extend the mentioned Gaussian-based filters by introducing the CSN distribution into the filtering problem.

A skewed version of the KF for linear systems was proposed by Naveau et al. (2005). They defined the filter in an extended state space model. Our proposed algorithms work for linear and nonlinear systems in a unified setting with structure similar to the KF and EnKF. Computational aspects are studied to handle the ensemble-based fitting and challenges related to the skewness dimension over many filtering time steps are discussed for all the KF variants.

In Section 2 we outline the modeling assumptions and present the CSN distribution. In Section 3 we present the CSNKF under linear modeling assumptions. In Section 4 we similarly present the CSNKF under nonlinear modeling assumptions. In Section 5 we illustrate the methodologies using numeric examples.

2. Background

2.1. Notation

Throughout this paper we use $\boldsymbol{x}_t \in \Re^{n_x \times 1}$ as a n_x dimensional distinction of interest at time $t = 1, \dots, T$. We assume that the dynamics of these state variables are represented by a set of difference equations with additive noise $\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_{t-1}) + \boldsymbol{\eta}_t$, where $\boldsymbol{f}(\cdot) : \Re^{n_x \times 1} \longmapsto \Re^{n_x \times 1}$ is a general linear or nonlinear function and $\boldsymbol{\eta}_t \in \Re^{n_x \times 1}$ is independent additive process noise with known distribution. If the system dynamics are linear, we use the notation $\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{\eta}_t$, where $\boldsymbol{F} \in \Re^{n_x \times n_x}$.

The notation $\boldsymbol{x} \sim \pi(\cdot)$ is used to show that variable $\boldsymbol{x}$ is distributed according to the probability density function (pdf) $\pi(\cdot)$. The Markov assumption about the states results in $\pi(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \dots \boldsymbol{x}_0) = \pi(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$, and $\pi(\boldsymbol{x}_T, \boldsymbol{x}_{T-1}, \dots \boldsymbol{x}_0) = \pi(\boldsymbol{x}_T|\boldsymbol{x}_{T-1}) \cdots \pi(\boldsymbol{x}_1|\boldsymbol{x}_0)\, \pi(\boldsymbol{x}_0)$, where $\pi(\boldsymbol{x}_0)$ is the initial condition distribution.

The observation equation is $\boldsymbol{d}_t = \boldsymbol{h}(\boldsymbol{x}_t) + \boldsymbol{\epsilon}_t$, where $\boldsymbol{d}_t \in \Re^{n_d \times 1}$ and $\boldsymbol{\epsilon}_t$ is the independent additive observation noise with known distribution. Thus, we assume that the data at different times are mutually independent given the state. The likelihood for the data is $\pi(\boldsymbol{d}_t|\boldsymbol{x}_t)$. The notation $\boldsymbol{D}_t = [\boldsymbol{d}_1, \boldsymbol{d}_2, \dots \boldsymbol{d}_t]$ is used for the collection of data from time 1 to t. Here we assume a linear or weakly nonlinear relationship between the observations and the state variables. Thus, we linearize the measurement equation using a first order Taylor series expansion to get $\boldsymbol{h}(\boldsymbol{x}_t) \approx \boldsymbol{h}_0 + \boldsymbol{H}\boldsymbol{x}_t$, where $\boldsymbol{H} \in \Re^{n_d \times n_x}$. Fig. 1 illustrates the modeling assumptions graphically. The presented filtering methods extend naturally to time-varying systems, i.e. when $\boldsymbol{f}$, $\boldsymbol{F}$, $\boldsymbol{h}$ or $\boldsymbol{H}$ change over time.

2.2. The filtering problem

Our goal is to assess the distribution of the state vector given all the currently available observations, i.e. the pdf $\pi(\boldsymbol{x}_t|\boldsymbol{D}_t)$ at each time index t. A recursive formulation gives the exact solution to this filtering problem by the following two steps. First, derive the one-step predictive pdf:

$$\pi(\boldsymbol{x}_t|\boldsymbol{D}_{t-1}) = \int \pi(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\, \pi(\boldsymbol{x}_{t-1}|\boldsymbol{D}_{t-1})\, d\boldsymbol{x}_{t-1}, \tag{1}$$