Automatica 85 (2017) 91–99

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Bayesian state estimation on finite horizons: The case of linear state–space model[☆]

Shunyi Zhao[a], Biao Huang[a], Yuriy S. Shmaliy[b]

[a] Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta T6G 2G6, Canada
[b] Department of Electronics Engineering, Universidad de Guanajuato, Salamanca 36885, Mexico

ARTICLE INFO

Article history:
Received 31 May 2016
Received in revised form 20 March 2017
Accepted 15 June 2017

Keywords:
State estimation
Finite impulse response
Optimal estimation
Maximum likelihood
Robustness

ABSTRACT

The finite impulse response (FIR) filter and infinite impulse response filter including the Kalman filter (KF) are generally considered as two different types of state estimation methods. In this paper, the sequential Bayesian philosophy is extended to a filter design using a fixed amount of most recent measurements, with the aim of exploiting the FIR structure and unifying some basic FIR filters with the KF. Specifically, the conditional mean and covariance of the posterior probability density functions are first derived to show the FIR counterpart of the KF. To remove the dependence on initial states, the corresponding likelihood is further maximized and realized iteratively. It shows that the maximum likelihood modification unifies the existing unbiased FIR filters by tuning a weighting matrix. Moreover, it converges to the Kalman estimate with the increase of horizon length, and can thus be considered as a link between the FIR filtering and the KF. Several important properties including stability and robustness against errors of noise statistics are illustrated. Finally, a moving target tracking example and an experiment with a three degrees-of-freedom helicopter system are introduced to demonstrate effectiveness.

1. Introduction

As one of the classical methods, the Bayesian filter provides a general solution to the state estimation problem, i.e., inferring a process $x_{n:0} := \{x_n, x_{n-1}, \ldots, x_0\}$ (known as states) from a set of observations $y_{n:0} := \{y_n, y_{n-1}, \ldots, y_0\}$. Let $p(x_n|y_{n:0})$ denote the probability density function (pdf) of x_n conditional on $y_{n:0}$, a posterior mean estimate is then given by

$$\hat{x}_{n|n} \triangleq \mathbb{E}[x_n|y_{n:0}] = \int x_n p(x_n|y_{n:0})\mathrm{d}x_n, \tag{1}$$

where $\mathbb{E}[x|y]$ is the expected value of x given y. If we assume that the states follow a hidden first-order Markov process with transition pdf $p(x_n|x_{n-1})$ and likelihood $p(y_n|x_n)$, the posterior pdf $p(x_n|y_{n:0})$ in (1) can be efficiently computed in a recursive manner with the well known recursions between *prediction* and *updating* (Anderson & Moore, 1979):

$$p(x_n|y_{n-1:0}) = \int p(x_n|x_{n-1})p(x_{n-1}|y_{n-1:0})\mathrm{d}x_{n-1}, \tag{2}$$

$$p(x_n|y_{0:n}) \propto p(y_n|x_n)p(x_n|y_{n-1:0}). \tag{3}$$

In this paper, we mainly focus on the linear Gaussian state–space model, where the Kalman filter (KF) provides a good interpretation of the sequential Bayesian filtering. For such a model, there is no filter other than the KF that provides optimal estimate in a simpler and faster way (Gelb, 1974). Consequently, numerous methods have been proposed to modify or extend the original algorithm to deal with some *unsatisfied* environments such as nonlinear models, missing measurements, uncertain parameters, and correlated noise sources (Abdallah, Gning, & Bonnifait, 2008; Cox, 1956; Karasalo & Hu, 2011; Simon, 2006; Yang & Yin, 2017), which have been proven to be effective in both theory and application. A common feature of these methods is that the current state is estimated based on all the past measurements. That is, we are interested in computing $p(x_n|y_{n:0})$. When model errors exist and when the filter operates over long time intervals, this operation may make estimates fall unacceptably away from the true values (Jazwinski, 1968). By analyzing the optimal filtering problem along the lines of the KF theory, Jazwinski concluded in Jazwinski (1970) that many flaws of the KF are due to this

☆ This work was supported by NSERC (Grant No.: IRCPJ 417793-10), AITF (Grant No.: PSI15502-2013). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wei Xing Zheng under the direction of Editor Torsten Söderström.

E-mail addresses: shunyi@ualberta.ca (S. Zhao), bhuang@ualberta.ca (B. Huang), shmaliy@ugto.mx (Y.S. Shmaliy).

http://dx.doi.org/10.1016/j.automatica.2017.07.043
0005-1098/

infinite impulse response (IIR) structure. Accordingly, another type of method named as finite impulse response (FIR) filter or limited memory filter has been proposed, which estimates states based *only* on a fixed amount of most recent data (Kwon & Han, 2006). In contrast to the IIR structure, the FIR counterpart exhibits many intrinsic advantages such as bounded-input bounded-output stability and better robustness against both numerical errors and temporal model uncertainties (Ahn, 2014; Kwon, Kim, & Han, 2002; Shmaliy, 2010, 2011; Zhao, Huang, Shmaliy, & Liu, 2015).

In spite of tremendous progress achieved to date, the FIR state estimation remains unconventional, and the gap between it and the well-established Kalman filtering still exists. The main reason we believe is that the Bayesian state estimation based on finite measurements has not been addressed systematically. Specifically, the existing FIR filters are obtained from different perspectives, and no unification study is available. Because of that, the FIR counterpart of KF has not been reported so far, and the existing unbiased FIR filters turn out to have completely different formulations (Kwon et al., 2002; Shmaliy, 2011; Zhao, Huang, et al., 2015; Zhao, Shmaliy, & Liu, 2015, 2016), which are not easy to be understood. The necessity of concise and explicit FIR filtering framework motivates our present study.

In this paper, we bridge the gap and extend the sequential Bayesian philosophy to a fixed amount of measurements in linear state–space systems. The main contributions of this paper are as follows. (1) As a FIR counterpart of the KF, the posterior pdf is obtained based on finite measurements and known initial distributions. (2) To remove the dependence on initial states of each estimation horizon, the batch maximum likelihood (ML) estimate is derived. It shows that the ML algorithm unifies the existing unbiased FIR filters (the unbiased FIR filter (Shmaliy, 2011) that omits noise statistics and the optimal unbiased solution proposed in Zhao et al. (2015)) with the weighted least square form. (3) With the concern of computational cost, an iterative ML realization is derived. Its difference from the KF in each iteration is shown analytically, and moreover, several useful properties including stability and improved robustness are demonstrated.

Notations: Throughout this paper, $\mathbb{R}^K$ denotes the K dimensional Euclidean space, $\mathbb{E}[\cdot]$ represents the statistical expectation, $I_{a\times a}$ and $0_{a\times a}$ refer to identity matrix and zero matrix of $a \times a$ dimensions, $\mathrm{tr}(\cdot)$ denotes the trace operator, $\mathcal{N}(x, P)$ is the Gaussian pdf with mean x and variance P, and $\mathrm{diag}(b_1 \cdots b_m)$ denotes a diagonal matrix with diagonal elements $b_1, \ldots, b_m$.

2. Preliminaries and problem formulation

We formulate the state estimation problem under the framework of a linear state–space model:

$$x_n = F_n x_{n-1} + G_n w_n, \tag{4}$$

$$y_n = H_n x_n + v_n, \tag{5}$$

where $x_n \in \mathbb{R}^\kappa$ is the state, $y_n \in \mathbb{R}^\rho$ is the measurement, $G_n \in \mathbb{R}^{\kappa\times\tau}$, $F_n \in \mathbb{R}^{\kappa\times\kappa}$ and $H_n \in \mathbb{R}^{\rho\times\kappa}$ are the system matrices, and $w_n \in \mathbb{R}^\tau$ and $v_n \in \mathbb{R}^\rho$ denote the process and measurement noises that are white Gaussian with zero mean and known covariance, i.e., $w_n \sim \mathcal{N}(0, Q_n)$ and $v_n \sim \mathcal{N}(0, R_n)$. It is assumed that w_n, v_n and $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$ are pairwise uncorrelated at each sampling instant. To construct a stable filter, $[F_n\ G_n Q_n^{1/2}]$ and $[F_n\ H_n]$ are assumed to be uniformly stabilizable and uniformly detectable, where $Q_n^{1/2}(Q_n^{1/2})^T = Q_n$ (Anderson & Moore, 1981).

With models (4) and (5), we can compute the posterior pdf $p(x_n|y_{n:0})$ using the KF, which is a full-information based strategy. If the model is perfect, this method is the best without doubt. Otherwise, the response to unpredictable dynamics may render an unrealistic small filter gain, by which the measurement information is unreasonably ignored, especially when a filter operates over a long time period. Defining a set of measurements $y_{n:m} := \{y_n, \ldots, y_m\}$, the problem considered can now be formulated as follows: Given the linear Gaussian state–space model (4) and (5), we show the FIR counterpart of the KF by calculating the posterior pdf in a finite horizon with the aim of giving an insight into the FIR structure. To remove the effect of initial values with respect to each estimation horizon, we further maximize the likelihood $p(y_{n:m}|x_n)$, and derive its fast prediction/correction formulation. Some important properties as well as the trade-off between the proposed ML FIR algorithm and the KF will also be illustrated.

3. FIR counterpart of the KF

In this section, the posterior pdf is calculated in a finite horizon with known initial distribution to show the relationships between the FIR methods and KF. Toward this end, the extended state–space model that resembles the original model (4) and (5) over the time interval $[m = n - N + 1, n]$ is constructed below, where N is the horizon length.

3.1. Extended state–space model

Using the forward-in-time solution and transforming all the state dynamic and measurement equations within $[m, n]$ with respect to the variable x_{m-1}, it is not difficult to find

$$X_{n,m} = F_{n,m} x_{m-1} + G_{n,m} W_{n,m}, \tag{6}$$

$$Y_{n,m} = H_{n,m} x_{m-1} + L_{n,m} W_{n,m} + V_{n,m}, \tag{7}$$

where $X_{n,m} = [x_n^T, x_{n-1}^T, \ldots, x_m^T]^T$, $W_{n,m} = [w_n^T, w_{n-1}^T, \ldots, w_m^T]^T$, $Y_{n,m} = [y_n^T, y_{n-1}^T, \ldots, y_m^T]^T$ and $V_{n,m} = [v_n^T, v_{n-1}^T, \ldots, v_m^T]^T$ are the extended vectors. The extended matrices $F_{n,m}$, $G_{n,m}$, $H_{n,m}$ and $L_{n,m}$ are given as, respectively,

$$F_{n,m} = \left[(\mathcal{F}_n^m)^T, (\mathcal{F}_{n-1}^m)^T, \ldots, F_m^T\right]^T,$$

$$G_{n,m} = \begin{bmatrix} G_n & F_n G_{n-1} & \cdots & \mathcal{F}_n^{m+2} G_{m+1} & \mathcal{F}_n^{m+1} G_m \\ 0 & G_{n-1} & \cdots & \mathcal{F}_{n-1}^{m+2} G_{m+1} & \mathcal{F}_{n-1}^{m+1} G_m \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & G_{m+1} & F_{m+1} G_m \\ 0 & 0 & \cdots & 0 & G_m \end{bmatrix},$$

$H_{n,m} = \bar{H}_{n,m} F_{n,m}$, and $L_{n,m} = \bar{H}_{n,m} G_{n,m}$ with $\bar{H}_{n,m} = \mathrm{diag}\big(H_n, H_{n-1}, \ldots, H_m\big)$ is a diagonal matrix and $\mathcal{F}_i^j = F_i F_{i-1} \cdots F_j$. Generally, $i \geqslant j$ and $\mathcal{F}_i^j = F_j$ when $i = j$. Note that if $N = n$, x_{m-1} becomes the initial state x_0, and the finite estimation horizon $[m, n]$ becomes the full observation interval $[1, n]$. Here, we exclude y_0 as $p(x_0)$ is known.

3.2. Posterior estimation

Using the extended state dynamic equation (6), the transition equation from x_{m-1} to x_n can be represented as

$$x_n = \mathcal{F}_n^m x_{m-1} + \mathcal{W}_{n,m}, \tag{8}$$

where $\mathcal{W}_{n,m} = \bar{G}_{n,m} W_{n,m}$, and $\bar{G}_{n,m} \triangleq [G_n\ F_n G_{n-1}\ \ldots\ \mathcal{F}_n^{m+2} G_{m+1}\ \mathcal{F}_n^{m+1} G_m]$ denotes the first row vector of $G_{n,m}$. By moving $G_{n,m} W_{n,m}$ from the right-hand side of (8) to the left-hand side, dividing both sides with $\mathcal{F}_n^m$, and substituting it into (7), the measurement equation becomes

$$Y_{n,m} = \tilde{H}_{n,m} x_n + \mathcal{L}_{n,m}, \tag{9}$$

where $\tilde{H}_{n,m} = H_{n,m}(\mathcal{F}_n^m)^{-1}$ and $\mathcal{L}_{n,m} = (L_{n,m} - \tilde{H}_{n,m}\bar{G}_{n,m})W_{n,m} + V_{n,m}$. Here, $\mathcal{F}_n^m$ is assumed to be invertible and we will show how to remove this assumption later. Since $W_{n,m}$ and $V_{n,m}$ are