Automatica 49 (2013) 816–820

Contents lists available at SciVerse ScienceDirect

Automatica

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

Brief paper

State estimation with partially observed inputs: A unified Kalman filtering approach☆

Baibing Li [1]

School of Business & Economics, Loughborough University, Loughborough LE11 3TU, United Kingdom

ARTICLE INFO

Article history:
Received 21 March 2012
Received in revised form
25 September 2012
Accepted 16 October 2012
Available online 18 January 2013

Keywords:
Bayesian inference
Data aggregation
Input observability
Kalman filters
State space models

ABSTRACT

For linear stochastic time-varying state space models with Gaussian noises, this paper investigates state estimation for the scenario where the input variables of the state equation are not fully observed but rather the input data are available only at an aggregate level. Unlike the existing filters for unknown inputs that are based on the approach of minimum-variance unbiased estimation, this paper does not impose the unbiasedness condition for state estimation; instead it incorporates a Bayesian approach to derive a modified Kalman filter by pooling the prior knowledge about the state vector at the aggregate level with the measurements on the output variables at the original level of interest. The estimated state vector is shown to be a minimum-mean-square-error estimator. The developed filter provides a unified approach to state estimation: it includes the existing filters obtained under two extreme scenarios as its special cases, i.e., the classical Kalman filter where all the inputs are observed and the filter for unknown inputs.

1. Introduction

State space modeling is widely used in various engineering fields. It also plays an important role in econometrics for time series analysis and forecasting (see, e.g. West & Harrison, 1997) with applications to economics, finance, and marketing, such as modeling arbitrage pricing and exchange rates (Priestley, 1996; Wells, 2010), and modeling sales growth and brand awareness (Wierenga, 2010). Recently it has also become a very popular approach to variable–coefficient regression modeling in econometrics.

In practice, modeling and decision-making depend on the availability of data measured on the variables of interest. The classical Kalman filter, a technique commonly used in state space models for rapidly updating the estimated state vector, considers an extreme scenario where all the input variables are observed. Recently, considerable attention has also been paid to the other extreme scenario where no input information is available: a set of recursive formulas has been derived via the approach of minimum-variance unbiased estimation. See Darouach and Zasadzinski (1997), Gillins and De Moor (2007) and Kitanidis (1987), among many others, for the recent development.

This paper complements the aforementioned methods and investigates state estimation when the input variables in state space models are not fully observed but rather they are available only at an aggregate level. This is a problem that has been recognized for a long time but has not yet been satisfactorily solved. In the literature there are three commonly used approaches: (a) the unobserved input variables are assumed to have little impact on the state variables so they are ignored; (b) an extra model is stipulated for the unobserved input variables; (c) the entire state space model is built at the aggregate level rather than at the level of interest.

For the first approach, the assumption that the unobserved input variables are ignorable may not be realistic in applications, and thus it can cause considerably large modeling errors. For traffic density estimation, for instance, Gazis and Liu (2003) assumed that lane changes of vehicles were not common and hence lane-change maneuvers, as the inputs of their state space model, were ignored. As a result, the modeling errors will become large for the roadways with substantial lane-changes.

With respect to the second approach, one commonly used method is to treat the unknown inputs as a stochastic process with a known description (known mean and covariance, for example) or as a constant bias (see, e.g. Friedland, 1969; Ignani, 1990; Zhou, Sun, Xi, & Zhang, 1993). Because more assumptions must be imposed, it is in general not an ideal solution when little is known about the input variables. For example, in the study of Australian state populations in Doran (1996), the net migration arrivals at the individual state level were treated as the input variables. These input variables, however, are not directly observed in non-census years. Doran (1996) assumed that they follow an AR(1) process and

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wolfgang Scherrer under the direction of Editor Torsten Söderström.

E-mail address: b.li2@lboro.ac.uk.

[1] Tel.: +44 1509 22884; fax: +44 1509 223960.

0005-1098/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2012.12.007

the coefficients of the AR(1) lags are common for all the states. Clearly, these assumptions are difficult to validate due to the lack of data. Kitanidis (1987) also discussed applications in geophysical and environmental fields where one cannot make any assumptions about the evolution of the unknown input variables.

Although the third approach is numerically feasible, Kalman filtering at an aggregate level may not be able to provide sufficient information for the problems under investigation. For instance, some existing studies in traffic studies (e.g., Wang & Papageorgiou, 2005) consider traffic modeling at an aggregate level, the segment level, where the traffic across all different lanes within a roadway segment was aggregated and modeled. Consequently, this approach is unable to provide lane-level traffic information which is crucial for some applications such as incident detection.

In this paper we assume that the unknown inputs at the level of interest have substantial impact on the system and they are not ignorable. In addition, we do not impose any extra assumptions on the unobserved inputs. Rather, we will develop a new method that makes use of the partially observed inputs to estimate the current state variables at the level of interest.

Although most of the recent studies on state estimation for unknown inputs have used the approach of minimum-variance unbiased estimation, we will incorporate a Bayesian approach in this paper. Bayesian analysis can be used to derive the classical Kalman filter (see, e.g. West & Harrison, 1997), and it is also a convenient method for generalizing the classical Kalman filter to solve more complicated nonlinear and/or non-Gaussian problems; see, e.g. the non-Gaussian Kalman filter in Li (2009) and the particle filter in Simon (2006).

We will show that Bayesian inference is a natural way to handle partially available input information. Unlike the existing studies (e.g. Darouach & Zasadzinski, 1997; Gillins & De Moor, 2007 and Kitanidis, 1987), there is no need to impose the unbiasedness condition in this paper. We show that under the assumption of Gaussian noise terms, the developed filter is optimal in the sense of minimum mean square error within the class of all estimators having a finite second moment. The Bayesian approach used in this paper neither makes any assumptions on the input variables nor directly estimates them at each time point (as done in Gillins and De Moor (2007), for instance). Instead, the prior knowledge about the state vector contained in the state equation is aggregated to the level at which the inputs are observed, which is then pooled with the current measurements on the outputs via Bayesian inference so that the estimated state vector is updated at each time step. We also show that the resulting recursive formulas provide a unified filtering approach for the problem where the availability of the input information ranges from all to nothing. In particular, it includes the classical Kalman filter (where all input variables are observed) and the filter for unknown inputs as its special cases.

2. Problem formulation and examples

2.1. Notation

Consider a linear discrete-time stochastic time-varying system in the form

$$x_{k+1} = A_k x_k + G_k d_k + w_k, \tag{1}$$

$$y_k = C_k x_k + v_k, \tag{2}$$

where $x_k = [x_{1,k}, \dots, x_{n,k}]^T \in R^n$ is the state vector, $d_k = [d_{1,k}, \dots, d_{m,k}]^T \in R^m$ is the input vector, and $y_k = [y_{1,k}, \dots, y_{p,k}]^T \in R^p$ is the measurement vector at each time step k. The process noise $w_k \in R^n$ and the measurement noise $v_k \in R^p$ are assumed to be mutually independent, and each follows a Gaussian distribution with zero mean and a known covariance matrix, $Q_k = E[w_k w_k^T] > 0$ and $R_k = E[v_k v_k^T] > 0$ respectively. Following the existing studies, we further assume that the initial state x_0 is independent of w_k and v_k with a known mean $\hat{x}_0$ and covariance matrix $P_0 > 0$.

We investigate the scenario where the input vector d_k is not fully observed at the level of interest. Instead some (or all) input data are available only at an aggregate level. Specifically, let D_k be a $q_k \times m$ known matrix with $0 \le q_k \le m$ and F_{0k} an orthogonal complement of D_k^T. It is assumed that the input data are available only on some linear combinations $D_k d_k$:

$$r_k = D_k d_k, \tag{3}$$

where r_k is observed at each time step k, and no information about $\delta_k = F_{0k}^T d_k$ is available. Hence, δ_k is assumed to have a noninformative distribution, i.e., it has a probability density function $f(\delta_k)$ for which all values of δ_k are equally likely to occur:

$$f(\delta_k) \propto 1. \tag{4}$$

The matrix D_k characterizes the availability of input information at each time step k. It includes two extreme scenarios that are of practical importance: (a) when $q_k = 0$, D_k is an empty matrix and thus no information on the inputs is available. This is the scenario investigated in Darouach and Zasadzinski (1997), Gillins and De Moor (2007) and Kitanidis (1987); (b) when $q_k = m$ and D_k is an identity matrix, it corresponds to the case that the complete input information is available. This is the case that the classical Kalman filter applies to. In some applications, the dimension q_k may vary from time to time. For instance, in economics and many other social sciences, the input data may be available at a microscopic level during census years but only at an aggregate level during non-census years.

To illustrate the scenario that input variables are partially observed, two examples are considered below, both involving a state equation of the following form:

$$x_{i,k+1} = x_{i,k} + \tilde{d}_{i,k} + u_{i,k} + w_{i,k}. \tag{5}$$

2.2. Estimation of Australian state populations

The study in Doran (1996) considered using the state space Eq. (5) to characterize the dynamic nature of the evolution of Australian state populations, where $x_{i,k}$ represents the population of state (or territory) i in year k, $u_{i,k}$ is the observed natural increase (births minus deaths), and $w_{i,k}$ is the corresponding error term. $\tilde{d}_{i,k}$ is the net migration arrivals in state i in year k. In non-census years the net migration arrivals are observed only at the national level. Hence, it is the linear combination of the net migration arrivals of the individual states, $\sum_{i=1}^{n} \tilde{d}_{i,k}$, that is available. Now we define the input variables to be $d_{i,k} = \tilde{d}_{i,k} + u_{i,k}$. So for this problem, the individual input variables $d_{i,k} (i = 1, \dots, n)$ in non-census years are observed only at an aggregate level (national level) and the matrix D_k in Eq. (3) is a row-vector of ones, whereas in census years all inputs $d_{i,k} (i = 1, \dots, n)$ are observed so D_k is an identity matrix.

2.3. Estimation of traffic densities

Intelligent transport systems for traffic surveillance require some fundamental information including traffic density. Traffic density is defined as the number of vehicles that occupy one unit length of road space per lane. Here we focus on a road segment with n lanes that is a detection zone with an upstream detector and a downstream detector at the entrance and exit of each lane respectively (see, e.g. Gazis & Liu, 2003). The two detectors count the vehicles passing through. See Li (2009) for a detailed description of the detectors.

The traffic conservation Eq. (5) is commonly used in the literature, where $x_{i,k}$ denotes the total number of vehicles in lane i at time step k, and $u_{i,k}$ represents the difference in the numbers