




This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.




CC creative commons
COMMONS DEED


Attribution-NonCommercial-NoDerivs 2.5


You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.


 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

State estimation with partially observed inputs: A unified Kalman filtering approach [★]

Baibing Li ^a

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

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 is 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.

Key words: Bayesian inference; Data aggregation; Input observability; Kalman filters; State space models.

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 and 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.

[★] This paper was not presented at any IFAC meeting. Corresponding author B. Li. Tel. +44-1509-228841. Fax +44-1509-223960.

Email address: b.li2@lboro.ac.uk (Baibing Li).

Friedland (1969); Ignani (1990); Zhou *et al.* (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 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 and 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 and 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 and Zasadzinski (1997); Gillins and 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 did 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

$$x_{k+1} = A_k x_k + G_k d_k + w_k, \quad (1)$$

$$y_k = C_k x_k + v_k, \quad (2)$$

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 $0 \leq q_k \leq m$ and F_{0k} an orthogonal complement of D_k^T . It is assumed that the input data is available only on some linear combinations $D_k d_k$:

$$r_k = D_k d_k, \quad (3)$$

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:

$$f(\delta_k) \propto 1. \quad (4)$$

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:

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

2.2 Estimation of Australian state populations

The study in Doran (1996) considered using the state space equation (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 equation (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 and Liu (2003)). The two detectors count the vehicles passing through. See Li (2009) for a detailed description of the detectors.

The traffic conservation equation (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 of vehicles that enter and leave the upstream and downstream detectors of lane i . The quantity $u_{i,k}$ is directly available from the detectors. However, usually no sensors are installed within a free-way segment. Hence, $\tilde{d}_{i,k}$, the vehicles' net gain due to lane-change maneuvers, is not observed in (5). We note, however, the net gain of lane-changing vehicles aggregated across all the lanes is equal to zero due to traffic conservation, i.e. $\sum_{i=1}^n \tilde{d}_{i,k} = 0$. Consequently, the input variables defined as $d_{i,k} = \tilde{d}_{i,k} + u_{i,k}$ are observable at the aggregate level (the segment level).

3 Main results

3.1 Modified Kalman filter

Consider state space model (1) and (2) with the input data available only at an aggregate level as specified in equation (3). Without loss of generality, it is assumed that the column-rank of D_k^T is equal to q_k ; otherwise the redundant rows of D_k can be eliminated.

In this paper we incorporate a Bayesian perspective and follow the approach in Li (2009) and West and Harrison (1997): at each time step k the state equation provides the prior knowledge about the current state vector, and the observation equation specifies the likelihood function. The updated estimate of the state vector is obtained from the posterior distribution.

Following the existing studies (e.g. Kitanidis (1987)), we restrict our interest to the case where G_k has a full-column rank, m . Let G_k^\perp denote an orthogonal complement of G_k and let $\Omega_k = [G_k, G_k^\perp]$. In numerical computation, we can compute a QR decomposition for G_k , $G_k = U_k \Psi_k$, where U_k is an orthogonal matrix, $\Psi_k = [\Lambda_k^T, O^T]^T$, and Λ_k is an upper triangular matrix. Then G_k^\perp can be chosen as $U_k [O^T, I^T]^T$. Define

$$\tilde{D}_k = \begin{pmatrix} D_k & O \\ O & I \end{pmatrix} \text{ and } \Pi_k = \begin{pmatrix} D_{k-1} \\ C_k G_{k-1} \end{pmatrix},$$

where I is an identity matrix and O is a matrix of zeros such that \tilde{D}_k is an $(q_k + (n - m)) \times n$ matrix. Let $M_k = \tilde{D}_k \Omega_k^{-1}$. The main result is summarized as follows.

Theorem 1 *For state space model (1) and (2) with $\text{rank}(G_k) = m$, suppose that the input data is available only at an aggregate level specified by equation (3), and the matrix Π_k has a full column-rank. Then the prior and posterior distributions for x_k at any time step k can be obtained sequentially as follows:*

(i) *Posterior of x_{k-1} for given $\{y_1, \dots, y_{k-1}\}$:*

$$x_{k-1} \sim N(\hat{x}_{k-1|k-1}, P_{k-1|k-1}).$$

(ii) *Prior for $s_k = M_{k-1} x_k$:*

$$s_k \sim N(M_{k-1} A_{k-1} \hat{x}_{k-1|k-1} + \tilde{r}_{k-1}, \tilde{P}_{k|k-1})$$

$$\text{with } \tilde{r}_k = [r_k^T, O^T]^T, \tilde{P}_{k|k-1} = M_{k-1} P_{k|k-1} M_{k-1}^T,$$

$$P_{k|k-1} = A_{k-1} P_{k-1|k-1} A_{k-1}^T + Q_{k-1}. \quad (6)$$

(iii) *Posterior of x_k for given $\{y_1, \dots, y_k\}$:*

$$x_k \sim N(\hat{x}_{k|k}, P_{k|k}),$$

where the posterior mean is given by

$$\begin{aligned} \hat{x}_{k|k} &= A_{k-1} \hat{x}_{k-1|k-1} + P_{k|k} M_{k-1}^T (M_{k-1} P_{k|k-1} \\ &\quad \times M_{k-1}^T)^{-1} \tilde{r}_{k-1} + K_k (y_k - C_k A_{k-1} \hat{x}_{k-1|k-1}), \end{aligned} \quad (7)$$

with the gain matrix

$$K_k = P_{k|k} C_k^T R_k^{-1}, \quad (8)$$

and the posterior variance is given by

$$P_{k|k} = [M_{k-1}^T (M_{k-1} P_{k|k-1} M_{k-1}^T)^{-1} M_{k-1} + C_k^T R_k^{-1} C_k]^{-1}. \quad (9)$$

To show Theorem 1, we need the following lemma. See the Appendix for proof.

Lemma 1 Consider linear model, $Y = X\beta + \xi$ with $\xi \sim N(0, V)$, where Y is an N -dimensional vector, X is an $N \times L$ matrix, and β is an L -dimensional vector of unknown parameters with a Gaussian prior distribution. β and ξ are independent of each other. $V > 0$ is a known covariance matrix. For two known matrices Γ and $\tilde{\Gamma}$ such that $[\Gamma^T, \tilde{\Gamma}^T]$ is an $L \times L$ invertible matrix, suppose that both the mean ϕ and covariance matrix $W > 0$ of $\gamma = \Gamma\beta$ are known, whereas $\lambda = \tilde{\Gamma}\beta$ follows $N(\mu, \tilde{W})$ with the hyperparameter vector μ having a noninformative prior. If the matrix $[\Gamma^T, X^T]^T$ has a full column-rank, then the posterior distribution of β given Y is $N(\theta, \Sigma)$, with posterior mean $\theta = \Sigma(\Gamma^T W^{-1} \phi + X^T V^{-1} Y)$ and posterior covariance matrix $\Sigma = (\Gamma^T W^{-1} \Gamma + X^T V^{-1} X)^{-1}$.

We now show Theorem 1 by induction. Assume the truth in part (i). We aggregate the state vector to the level at which the inputs are observed by defining $s_k = M_{k-1} x_k$. Noting $M_{k-1} G_{k-1} d_{k-1} = \tilde{r}_{k-1}$, we obtain the prior in part (ii) by combining equation (1) with part (i).

Next we show part (iii). Define $\tilde{M}_k = [F_{0k}^T, O] \Omega_k^{-1}$ and $\tilde{s}_k = \tilde{M}_{k-1} x_k$. We note that from part (ii), s_k follows a Gaussian distribution with a known mean and covariance matrix, whereas \tilde{s}_k has a Gaussian distribution with an unknown mean $M_{k-1} A_{k-1} \hat{x}_{k-1|k-1} + \delta_{k-1}$. From (4) and applying Lemma 1, we obtain the posterior of x_k which is Gaussian with mean

$$\hat{x}_{k|k} = P_{k|k} \{ M_{k-1}^T (M_{k-1} P_{k|k-1} M_{k-1}^T)^{-1} [M_{k-1} A_{k-1} \hat{x}_{k-1|k-1} + \tilde{r}_{k-1}] + C_k^T R_k^{-1} y_k \}.$$

The above equation can be re-arranged to have the form of equation (7). The covariance matrix in equation (9) is the result of applying Lemma 1.

In addition, let S_k denote $\begin{pmatrix} M_{k-1} \\ C_k \end{pmatrix}$. Under the conditions $Q_k > 0$ and $R_k > 0$, we obtain that $P_{k|k}$ is non-singular if S_k has a full column-rank. Since

$$S_k = \begin{pmatrix} D_{k-1} & O \\ O & I \\ C_k G_{k-1} & C_k G_{k-1}^+ \end{pmatrix} \Omega_{k-1}^{-1},$$

the condition that S_k has a full column-rank is equivalent to the condition that Π_k has a full column-rank.

The inductive proof is completed by noting that the results are true for $k = 1$ where $x_0 \sim N(\hat{x}_{0|0}, P_{0|0})$ with $\hat{x}_{0|0} = \hat{x}_0$ and $P_{0|0} = P_0$ respectively. \square

Remarks:

- (i) Collecting recursive formulas (6)-(9) gives a modified Kalman filter for the problem that the input data is available only at an aggregate level (3).
- (ii) The essence of the Bayesian approach here is pooling of the prior knowledge about the state vector at the aggregate level in part (ii) with the current observation (2) at the original level of interest so that the posterior mean $\hat{x}_{k|k}$ and posterior covariance matrix $P_{k|k}$ are computable.
- (iii) The condition that matrix Π_k has a full column-rank has a clear physical meaning: it ensures that the state vector is estimable at each time k .

Next, we turn to consider the computational issue. Equation (9) involves the inverse of an $n \times n$ matrix. When n is large, the computational cost is high. For the case that C_k has a full column-rank, the covariance matrix $P_{k|k}$ can be written as

$$P_{k|k} = T_k - T_k M_{k-1}^T [M_{k-1} (P_{k|k-1} + T_k) M_{k-1}^T]^{-1} M_{k-1} T_k, \quad (10)$$

with $T_k = (C_k^T R_k^{-1} C_k)^{-1}$. Clearly when q_k is small, the computational cost is reduced by using equation (10). This approach is particularly useful when both C_k and R_k are time-invariant so that T_k can be computed in advance. When C_k does not have a full column-rank, however, we need another approach.

Lemma 2 Let $P > 0$ and $R > 0$. Let F be an orthogonal complement of D^T such that $DF = O$. Suppose that both the matrix D^T and matrix $[C^T, D^T]^T$ have a full column-rank. Then we have

- (i) $[D^T (DPD^T)^{-1} D + C^T R^{-1} C]^{-1}$
 $= P - PC^T H^{-1} CP + [F - PC^T H^{-1} CF]$
 $\times [F^T C^T H^{-1} CF]^{-1} [F - PC^T H^{-1} CF]^T;$
- (ii) $[D^T (DPD^T)^{-1} D + C^T R^{-1} C]^{-1} C^T R^{-1}$
 $= PC^T H^{-1} + [F - PC^T H^{-1} CF]$
 $\times [F^T C^T H^{-1} CF]^{-1} F^T C^T H^{-1},$
where $H = CPC^T + R$.

See the Appendix for proof. Lemma 2 provides an alternative for the computation of the modified Kalman filter. Specifically, let F_k denote an orthogonal complement of M_k^T , where F_k can be constructed as follows: $F_k = \Omega_k \begin{pmatrix} F_{0k} \\ O \end{pmatrix}$. From Lemma 2, equations (8) and (9) can be computed as follows:

$$K_k = P_{k|k-1} C_k^T H_k^{-1} + [F_{k-1} - P_{k|k-1} C_k^T H_k^{-1} C_k F_{k-1}] \times [F_{k-1}^T C_k^T H_k^{-1} C_k F_{k-1}]^{-1} F_{k-1}^T C_k^T H_k^{-1}, \quad (11)$$

$$P_{k|k} = P_{k|k-1} - P_{k|k-1}C_k^T H_k^{-1} C_k P_{k|k-1} + [F_{k-1} - P_{k|k-1}C_k^T H_k^{-1} C_k F_{k-1}][F_{k-1}^T C_k^T H_k^{-1} C_k F_{k-1}]^{-1} \times [F_{k-1} - P_{k|k-1}C_k^T H_k^{-1} C_k F_{k-1}]^T, \quad (12)$$

with $H_k = C_k P_{k|k-1} C_k^T + R_k$.

We now turn to the assessment of the quality of the proposed filter. It is known that the filter developed in Kitanidis (1987) is optimal in the sense of unbiasedness and minimum variance within the class of all linear estimators. In Bayesian statistics, however, the concept of minimum-variance unbiased estimator does not apply because the parameters of interest are considered to be random variables rather than some fixed values. Instead, the quality of an estimator can be measured via the minimum-mean-square-error (MMSE) criterion.

Lemma 3 *Under the assumptions of Lemma 1, $\hat{\theta}(Y) = \Sigma(\Gamma^T W^{-1} \phi + X^T V^{-1} Y)$ is the unique MMSE estimator of β , where $\Sigma = (\Gamma^T W^{-1} \Gamma + X^T V^{-1} X)^{-1}$.*

The proof is immediate from Lehmann and Casella (1998) (Theorem 4.1.1, and Corollaries 4.1.2 & 4.1.4). Applying Lemma 3, we obtain that the proposed filter is an MMSE estimator. Note that under the assumption of Gaussian noise terms in (1) and (2), this result holds not only for the class of all linear estimators and for the class of all recursive estimators, but also true for the class of all estimators having a finite second moment.

3.2 A unified approach to state estimation

For the scenario where no input information is available, the filtering problem has recently been investigated by a number of researchers using the approach of minimum-variance unbiased estimation. For instance, the recursive formulas derived by Kitanidis (1987) are

$$P_{k|k-1} = A_{k-1} P_{k-1|k-1} A_{k-1}^T + Q_{k-1}, \quad (13)$$

$$\hat{x}_{k|k} = A_{k-1} \hat{x}_{k-1|k-1} + K_k (y_k - C_k A_{k-1} \hat{x}_{k-1|k-1}), \quad (14)$$

$$K_k = P_{k|k-1} C_k^T H_k^{-1} + [G_{k-1} - P_{k|k-1} C_k^T H_k^{-1} C_k G_{k-1}] \times [G_{k-1}^T C_k^T H_k^{-1} C_k G_{k-1}]^{-1} G_{k-1}^T C_k^T H_k^{-1}, \quad (15)$$

$$P_{k|k} = P_{k|k-1} - P_{k|k-1} C_k^T H_k^{-1} C_k P_{k|k-1} + [G_{k-1} - P_{k|k-1} C_k^T H_k^{-1} C_k G_{k-1}][G_{k-1}^T C_k^T H_k^{-1} C_k G_{k-1}]^{-1} \times [G_{k-1} - P_{k|k-1} C_k^T H_k^{-1} C_k G_{k-1}]^T, \quad (16)$$

with $H_k = C_k P_{k|k-1} C_k^T + R_k$.

It is of particular interest to ask in which way that the filter developed in this paper is linked to the existing filters. The answer to this question is summarized in the following theorem.

Theorem 2 *The recursive formulas (6)-(9) reduce to: (a) the classical Kalman filter when all the entries of the input vector d_k are observed; and (b) the filtering equations (13)-(16) when no information on d_k is available.*

Proof. First, when all the input variables are observed at the level of interest, D_k becomes an $m \times m$ identity matrix, and thus \tilde{D}_k is an $n \times n$ identity matrix. Hence, the orthogonal complement F_k of M_k^T is a zero-by-zero empty matrix. Consequently the last term on the right-hand-side of equations (11) and (12) vanishes, and equations (11) and (12) reduce to

$$K_k = P_{k|k-1} C_k^T H_k^{-1},$$

$$P_{k|k} = P_{k|k-1} - K_k C_k P_{k|k-1}.$$

Noting $\tilde{r}_{k-1} = G_{k-1} d_{k-1}$, equation (7) becomes

$$\hat{x}_{k|k} = A_{k-1} \hat{x}_{k-1|k-1} + G_{k-1} d_{k-1} + K_k [y_k - C_k (A_{k-1} \hat{x}_{k-1|k-1} + G_{k-1} d_{k-1})].$$

These recursive formulas are identical to the classical Kalman filter equations (see, e.g. Simon (2006)).

Next, we turn to the scenario that no input information is available. Since D_k in equation (3) reduces to a zero-by-zero empty matrix in this case, we have $\tilde{D}_k = [O, I]$. We choose $G_k^\perp = U_k [O, I]^T$, where U_k is the orthogonal matrix in the QR decomposition $G_k = U_k \Psi_k$ as defined before. Then we obtain $M_k = \tilde{D}_k \Omega_k^{-1} = [O, I] U_k^T$. Therefore, we can choose an orthogonal complement of M_k^T as $F_k = G_k$. We then obtain that equations (11) and (12) reduce to (15) and (16). Finally, equation (13) is immediate by noting $\tilde{r}_k = M_k G_k d_k = 0$. \square

4 Conclusions

We have investigated the state estimation problem where the input variables in the state space model are not fully observed but rather they are available only at an aggregate level. A Bayesian approach is used to derive a modified Kalman filter by pooling the prior knowledge about the state vector at the aggregate level with the measurements provided by the observation equation at the original level of interest. As a result, the unobserved input variables are taken out via data aggregation. This method can be applied to estimate the state vector at the original level of interest without imposing any further assumptions on the unobserved input variables.

Unlike many existing filters for unknown inputs, the filter developed in this paper is an MMSE estimator. Under the assumption of Gaussian noise terms, it can be obtained without imposing the unbiasedness condition and it is shown to be optimal within the class of all estimators having a finite second moment.

The developed filter provides a unified approach to state estimation where the availability of input information ranges from all to nothing. In particular, it includes the classical Kalman filter (where all inputs are

observed) and the filter for unknown inputs as its special cases.

Appendix. Proofs of lemmas

4.1 Proof of Lemma 1

Let $f(\mu)$, $f(\gamma)$ and $f(\lambda)$ be the marginal probability density functions of μ , γ , and λ respectively, and let $f(Y|\beta)$ be the likelihood function. Note that for given μ , γ and λ follow $N\left(\begin{pmatrix} \phi \\ \mu \end{pmatrix}, \begin{pmatrix} W & \tilde{C} \\ \tilde{C}^T & \tilde{W} \end{pmatrix}\right)$, where \tilde{C} denotes the covariance matrix. Since $f(\mu) \propto 1$, we obtain $f(\gamma, \lambda) = \int f(\gamma, \lambda|\mu)f(\mu)d\mu \propto \int f(\gamma, \lambda|\mu)d\mu = f(\gamma)$, where the final equality can be obtained by direct calculation of the integral. By Bayes' rule, the posterior distribution $f(\beta|Y)$ is

$$\begin{aligned} f(\beta|Y) &\propto f(Y|\beta)f(\beta) \propto f(Y|\beta)f(\gamma) \\ &\propto \exp\{-(Y - X\beta)^T V^{-1}(Y - X\beta)/2 \\ &\quad - (\Gamma\beta - \phi)^T W^{-1}(\Gamma\beta - \phi)/2\}. \end{aligned}$$

By completing the square on β , the exponent can be rewritten as $-(\beta - \theta)^T \Sigma^{-1}(\beta - \theta)/2 + \text{constant}$. This shows that the posterior distribution is a Gaussian distribution with mean θ and covariance matrix Σ . \square

4.2 Proof of Lemma 2

To show (i), we note that the left-hand-side of the equation in part (i) can be rewritten as

$$\begin{aligned} \{[P^{-1} + C^T R^{-1} C] - P^{-1/2}[I - P^{1/2} D^T (D P D^T)^{-1} \\ \times D P^{1/2}] P^{-1/2}\}^{-1}. \end{aligned}$$

The matrix $I - P^{1/2} D^T (D P D^T)^{-1} D P^{1/2}$ is the orthogonal projection matrix onto an orthogonal complement of the column space of $P^{1/2} D^T$. Now taking this orthogonal complement as $P^{-1/2} F$, we obtain that

$$\begin{aligned} I - P^{1/2} D^T (D P D^T)^{-1} D P^{1/2} \\ = P^{-1/2} F (F^T P^{-1} F)^{-1} F^T P^{-1/2}. \end{aligned}$$

So the left-hand-side of the equation in part (i) becomes

$$\{[P^{-1} + C^T R^{-1} C] - P^{-1} F (F^T P^{-1} F)^{-1} F^T P^{-1}\}^{-1}.$$

It then follows by applying the matrix inversion lemma (see, e.g. Simon (2006), pp12) twice, first to the whole equation, and then to $[P^{-1} + C^T R^{-1} C]^{-1}$. The proof for part (ii) is similar after inserting the result in part (i) into the left-hand-side of the equation in part (ii). \square

Acknowledgements

The author thanks the reviewers and the associate editor for their helpful comments on the earlier versions of this paper.

References

- Darouach, M., & Zasadzinski, M. (1997). Unbiased minimum variance estimation for systems with unknown exogenous inputs. *Automatica* 33, 717–719.
- Doran, H. E. (1996). Estimation under exact linear time-varying constraints with applications to population projections. *Journal of Forecasting* 15, 527–541.
- Friedland, B. (1969). Treatment of bias in recursive filtering. *IEEE Trans. Autom. Control AC-14*, 359–367.
- Gazis, D., & Liu, C. (2003). Kalman filtering estimation of traffic counts for two network links in tandem. *Transportation Research Part B* 37, 737–745.
- Gillins, S. & De Moor, B. (2007). Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica* 43, 111–116.
- Ignani, M. B. (1990). Separate bias Kalman estimator with bias state noise. *IEEE Trans. Autom. Control AC-35*, 338–341.
- Kitanidis, P. K. (1987). Unbiased minimum-variance linear state estimation. *Automatica* 23, 775–778.
- Lehmann, E. L., & Casella, G. (1998). *Theory of Point Estimation*. (2nd ed.). New York: Springer.
- Li, B. (2009). A non-Gaussian Kalman filter with application to the estimation of vehicular speed. *Technometrics* 51, 162–172.
- Priestley, R. (1996). The arbitrage pricing theory, macroeconomic and financial factors, and expectations generating processes. *Journal of Banking & Finance* 20, 869–890.
- Simon, D. (2006). *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. New York: Wiley.
- Wang, Y., & Papageorgiou, M. (2005). Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. *Transportation Research Part B* 39, 141–167.
- Wells, C. (2010). *The Kalman Filter in Finance*. London: Kluwer.
- West, M., & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. (2nd ed.). New York: Springer.
- Wierenga, B. (2010). *Handbook of marketing decision models*. New York: Springer.
- Zhou, D. H., Sun, Y. X., Xi, Y. G., & Zhang, Z. J. (1993). Extension of Friedland's separate bias estimation to randomly time-variant bias of nonlinear systems. *IEEE Trans. Autom. Control AC-38*, 1270–1273.