State Estimation for Systems with Unknown Inputs
Based on Variational Bayes Method

Junlong Sun  Jie Zhou†
College of Mathematics
Sichuan University
Chengdu, Sichuan 610064, China
junlong.sun@gmail.com, jzhou@scu.edu.cn

X. Rong Li
Department of Electrical Engineering
University of New Orleans
New Orleans, LA 70148, USA
xli@uno.edu

Abstract—This paper considers a probabilistic approach to state estimation for discrete-time dynamic systems with unknown inputs. A variational Bayes method is proposed to approximate the marginal posterior distributions of system state and input. In order to reduce the computational complexity, the complete-data likelihoods of system from the exponential family are considered, and the conjugate prior distributions are used to quantify the input. Then variational Bayesian learning procedures are derived to optimize the marginal distributions of the state and input. Specifically, recursive filtering for a linear Gaussian system is presented. As applications, state estimation for several important practical systems with unknown inputs is discussed. Related numerical simulations are provided to demonstrate the performance of the proposed method.

Keywords: State estimation, unknown inputs, Bayesian filtering, variational Bayes method.

I. INTRODUCTION

Optimal filtering theory for dynamic systems has been widely applied in many engineering and scientific fields [1], [4]. The problem of tracking moving targets in multisensor systems have received increasing attentions in the past three decades (see, e.g., [2], [3]). In many practical systems, model disturbances in the systems are often modeled as unknown input. Inappropriate dealing with the input may lead to degraded estimation performance or unreliable filtering and tracking results. For the linear Gaussian dynamic system, many of the techniques have been developed for linear state estimation with unknown input (see, e.g., [9], [10], [15], [19] and references therein). In some practical cases, state space models are neither Gaussian nor linear, so the aforementioned estimation techniques cannot be applied directly. Some work exists that is efficient for separate estimation of the unknown input for nonlinear dynamic systems (see, e.g., [8], [29]).

The Bayesian approach is a general method for state estimation. It is well-known that the Kalman filter is a special case of the Bayes filter, which is, however, often computationally complicated [6]. One of its approximation methods is Markov Chain Monte Carlo (MCMC) sampling, which uses numerical integration to side-step the problem of multiple integration.

However, the MCMC method is time-consuming and cannot ensure convergence [7]. Some approximation techniques for nonlinear filtering problems, especially those in the context of maneuvering target tracking were discussed in [22], [23].

The variational Bayes (VB) method applies variational ideas from the analysis of intractable latent variable models and approximates the posterior at a low computational cost [6], [7], [13]. This effective approach is becoming a useful tool for deterministic approximation for Bayesian inference. Some researchers have developed this method for linear dynamic systems (see, e.g., [5]–[7], [25], [31]). In [14], the variational learning method was adopted to solve jumping state-space systems. For a linear dynamic system with unknown statistical properties of measurement noise, the VB method was proposed in [24] for state estimation and generalized to multi-sensor systems in [12]. Recently, in [26], the VB approach was considered for state estimation where a dynamic model of unknown input is available. In [21], approaches to state estimation with unknown input for maneuvering target tracking were surveyed.

In this paper, we derive a probabilistic approach to estimate the state of a discrete-time dynamic system with an unknown input, in which no dynamic models of the input are assumed. In order to overcome the difficulties owing to the computational complexity of the fully Bayesian filtering for the dynamic systems, we assume that the complete-data likelihood distribution belongs to the exponential family, and then present an algorithm to learn the structure of hidden variables and approximate the joint posterior distribution of the system state and input. Moreover, we derive a recursive filtering algorithm with suitable priors based on the VB approach. VB treatment for a general linear dynamic system is presented, and the relation with the conjugate exponential family is discussed. As applications, some important practical systems with unknown inputs are discussed, including linear and nonlinear Gaussian state space models, mixture Gaussian noise, cluttered measurement models and sensor registration systems. Numerical experiments show that the proposed method has good state estimation performance for the state-space systems with uncertainties.

This paper is organized as follows. The problem of state estimation with unknown input is formulated in Section II.
In Section III, we briefly review the conjugate priors of the exponential family, and then apply the VB method to approximate the joint posterior distribution of the system state and input. Section IV provides analytical results of VB method for a linear dynamic system. Several numerical experiments to show the performance of the proposed method are presented in Section V. Section VI gives conclusions.

II. PROBLEM FORMULATION

Consider the following discrete-time dynamic system:

\[ x_{k+1} = f(x_k, r_k, \omega_k, k), \quad y_k = g(x_k, r_k, \nu_k, k), \]

where, \( k \) is the time index, \( x_k \in \mathbb{R}^n \) is the system state, \( y_k \in \mathbb{R}^p \) is the measurement, \( f(\cdot) \) and \( g(\cdot) \) respectively represent the state transition function and the measurement function, \( r_k \in \mathbb{R}^q \) is the unknown input, and \( \omega_k \) and \( \nu_k \) are process noise and measurement noise respectively.

The dynamic system (1) can also be expressed in probablistic terms:

\[ x_{k+1} \sim p(x_{k+1}|x_k, r_k), \quad y_k \sim p(y_k|x_k, r_k), \]

where, \( p(x_{k+1}|x_k, r_k) \) characterizes the state transition, and \( p(y_k|x_k, r_k) \) describes the measurement model. The initial density of the state is denoted by \( p(x_0) \). Once the measurements \( y_{1:k} = [y_1, \ldots, y_k]^T \) are given, we can use the Bayes filtering technique to predict and estimate the system state.

The main part of the Bayes filter includes computation of the marginal distribution of joint posterior \( p(x_k, r_k|y_{1:k}) \). Formally, a recursive solution to this problem consists of the following steps:

1) Initialization: the recursion starts from the initial distribution \( p(x_0) \);
2) Distribution prediction: calculate the joint prior distribution:

\[
p(x_k, r_k|y_{1:k-1}) = \int p(x_k|x_{k-1}, r_{k-1}) \cdot p(x_{k-1}, r_{k-1}|y_{1:k-1}) dx_{k-1} dr_{k-1};
\]

3) Distribution estimation: given the measurement \( y_k \), the predictive distribution is updated to the posterior distribution by Bayes’ rule:

\[
p(x_k, r_k|y_{1:k}) = \frac{p(y_k|x_k, r_k)p(x_k, r_k|y_{1:k-1})}{Z_k},
\]

where \( Z_k = \int p(y_k|x_k, r_k)p(x_k, r_k|y_{1:k-1})dx_k dr_k \) is the normalization constant.

In many cases, it is difficult to obtain the exact analytical solution of (3) [6], [13]. The objective of this paper is to derive a recursive VB filtering algorithm, which approximates joint posterior distribution, and to obtain the marginal posterior of the state and input effectively.

III. VARIATIONAL BAYESIAN FILTERING

In this section, we first review the conjugate exponential family, and then apply the results to quantify the uncertainties in the dynamic system using the VB method.

A. Conjugate Priors of Exponential Family

Consider the exponential family of distribution for a given parameter \( \eta \):

\[ p(x|\eta) = h(x) \exp(\phi(\eta)^T \cdot u(x) - \psi(\eta)), \quad (4) \]

where \( h(x) \) and \( u(x) \) are known functions, \( \phi(\eta) \) is the vector of natural parameters, and \( \psi(\eta) \) can be interpreted as the normalization constant

\[ \psi(\eta) = \ln \left( \int h(x) \exp(\phi(\eta)^T \cdot u(x))dx \right). \]

Assume the prior of \( \eta \) is

\[ p(\eta; \alpha^-, \beta^-) = \exp(\phi(\eta)^T \cdot \alpha^- - \psi(\eta, \beta^-) - H(\alpha^-, \beta^-)), \quad (5) \]

where \( \alpha^- \) and \( \beta^- \) are hyper-parameters of the prior, and \( H(\alpha^-, \beta^-) \) is the normalization constant. Applying Bayes’ rule, we have the posterior density (see [27]):

\[ p(\eta; \alpha, \beta) \propto \exp(\phi(\eta)^T \cdot \alpha - \tilde{\psi}(\eta, \beta)), \]

where \( \alpha \) and \( \beta \) are hyper-parameters

\[ \alpha = \alpha^- + y, \quad \beta = \beta^- + I, \quad (6) \]

\( y \) is the observation, and \( I \) is the identity matrix.

The prior density \( p(\eta; \alpha^-, \beta^-) \) and the posterior density \( p(\eta; \alpha, \beta) \) have the same form, and thus the Bayesian inference reduces to hyper-parameter update, where (5) is defined as conjugate prior to exponential family (4). Note that it holds regardless of the particular form of the complete-data likelihood distribution.

B. Variational Bayesian Treatment

Given measurements \( y_{1:k-1} \), the posteriors \( p(x_{k-1}|y_{1:k-1}) \) and \( p(r_{k-1}|y_{1:k-1}) \), we consider dynamic system (1) and assume the complete-data likelihood from the general exponential family

\[ p(y_k, x_k|r_k) = f_k(x_k, y_k) \exp(\phi(r_k)^T \cdot u_k(x_k, y_k) - \psi(r_k)). \]

(7)

To quantify the unknown input, assume that \( r_k \) has a prior density that is conjugate to (7)

\[ p(r_k; \alpha_k^-, \beta_k^-) = \exp(\alpha_k^- \cdot r_k - \tilde{\psi}(r_k, \beta_k^-) - H(\alpha_k^-, \beta_k^-)). \]

(8)

Then applying Bayes’ rule, we have

\[ p(r_k|x_k, y_{1:k}) \propto p(y_k, x_k|r_k, y_{1:k-1})p(r_k; \alpha_k^-, \beta_k^-) \]

(9)

Note that the analytic solution to (9) would be infeasible to evaluate or difficult to compute [6], [13]. Without computing the marginal posterior, we use the VB method to derive analytical approximations with a product of tractable marginal posteriors [6], [7], [25]. We assume the variables are conditionally independent, and factorize the joint posterior with two analytically manageable distributions:

\[ p(x_k, r_k|y_{1:k}) \approx Q(x_k, r_k) = Q_x(x_k)Q_r(r_k), \]

984
Specifically, the VB marginal distributions \( Q_{z}(x_k), Q_{r}(r_k) \), and \( Q(x_k, r_k) \) are some marginal and joint densities of \( x_k \) and \( r_k \). Applying Jensen’s inequality [18], we have

\[
\ln p(y_k | y_{1:k-1}) = \ln \int Q(x_k, r_k) \frac{p(x_k, r_k, y_k | y_{1:k-1})}{Q(x_k, r_k)} dx_k dr_k \\
\geq \int Q(x_k, r_k) \ln \left( \frac{p(x_k, r_k, y_k | y_{1:k-1})}{Q(x_k, r_k)} \right) dx_k dr_k \\
= \ln p(y_k | y_{1:k-1}) - \text{KL}(Q(x_k, r_k) || p(x_k, r_k | y_{1:k})) \\
\equiv J(Q(x_k, r_k)),
\]

where \( \text{KL}(p || q) \) is the Kullback-Leibler (KL) divergence from \( p \) to \( q \). Therefore, maximizing the variational lower bound \( J(Q(x_k, r_k)) \) is equivalent to minimizing the KL divergence from the approximating posterior \( Q(x_k, r_k) \) to the exact posterior \( p(x_k, r_k | y_{1:k}) \). It is clear

\[
\text{KL}(Q(x_k, r_k) || p(x_k, r_k | y_{1:k})) = \int Q(x_k, r_k) \ln \left( \frac{Q(x_k, r_k)}{p(x_k, r_k | y_{1:k})} \right) dx_k dr_k \\
= \int Q_z(x_k) Q_r(r_k) \ln \left( \frac{Q_z(x_k) Q_r(r_k)}{p(x_k, r_k | y_{1:k})} \right) dx_k dr_k.
\]

Given the measurement \( y_k \), we can minimize the KL divergence (10) with each of the factorized densities \( Q_z(x_k) \) and \( Q_r(r_k) \) in turn, while keeping the other fixed [6], [7]. Specifically, the VB marginal distributions \( Q_z(x_k) \) and \( Q_r(r_k) \) can be given as

\[
\ln Q_z(x_k) = \langle \ln p(x_k, r_k, y_k | y_{1:k-1}) \rangle_{r_k} + \text{const}, \quad (11) \\
\ln Q_r(r_k) = \langle \ln p(x_k, r_k, y_k | y_{1:k-1}) \rangle_{x_k} + \text{const}, \quad (12)
\]

where \( \langle \cdot \rangle_{x_k} \) and \( \langle \cdot \rangle_{r_k} \) denote the expectations with respect to \( Q_z(x_k) \) and \( Q_r(r_k) \), respectively. Applying (6), \( Q_z(x_k) \) and \( Q_r(r_k) \) can be obtained from the following procedure:

1) Optimize \( Q_r(r_k) \) for fixed \( Q_z(x_k) \):

\[
Q_r(r_k) \propto \exp \left( \phi(r_k)^T \cdot \alpha_k - \psi(r_k), \beta_k) \right), \quad (13)
\]

where \( \alpha_k \) and \( \beta_k \) are the hyper-parameters, and

\[
\alpha_k = \alpha_k^- + \langle u_k(x_k, y_k) \rangle_{x_k}, \quad \beta_k = \beta_k^+ + \beta_k^- + I.
\]

2) Optimize \( Q_z(x_k) \) for fixed \( Q_r(r_k) \):

\[
Q_z(x_k) = f_k(x_k, y_k) z_k(r_k, y_k) \\
\cdot \exp \left( \left( \phi(r_k)^T \cdot u_k(x_k, y_k) - \psi(r_k) \right) r_k \right), \quad (14)
\]

where \( z_k(r_k, y_k) \) is a normalization constant satisfying

\[
z_k(r_k, y_k)^{-1} = \int f_k(x_k, y_k) \\
\cdot \exp \left( \left( \phi(r_k)^T \cdot u_k(x_k, y_k) - \psi(r_k) \right) r_k \right) dx_k.
\]

Remark 3.1: According to the Darmois-Koopman-Pitman theorem [11], among all families of smooth and nowhere vanishing probability distributions whose domain does not vary with the parameter to be estimated, the exponential family is the only one that has a sufficient statistic whose dimension remains bounded as sample size increases. Therefore, the complete-data likelihood (7) is a reasonable assumption for deriving a recursive closed form for VB filtering.

Remark 3.2: The VB learning procedure (13) and (14) hold regardless of the particular observation distribution and always form an iteratively learning algorithm, which have several desirable properties including convergence and asymptotic normality [28].

Remark 3.3: If the measurement includes some transformation of the input, (13) and (14) can still hold for the transformed input. Under this circumstance, conjugate prior for the original input may not have the form (8) when the transformation is not invertible (see, e.g., the linear dynamic system discussed in Section IV). In this case, (11) and (12) can be employed to obtain the VB marginal distributions.

IV. LINEAR DYNAMIC SYSTEMS

In this section, we consider the following linear Gaussian dynamic system with unknown input:

\[
x_{k+1} = A_k x_k + F_k r_k + \omega_k, \\
y_k = H_k x_k + G_k r_k + \nu_k.
\]

Here \( A_k, F_k, H_k \) and \( G_k \) are known matrices with appropriate dimensions, and the process noise \( \omega_k \) and measurement noise \( \nu_k \) are Gaussian random variables with zero-mean and time-variaying covariances:

\[
E\{\omega_k \omega_k^T\} = Q_k \delta_{kl}, \quad E\{\nu_k \nu_k^T\} = R_k \delta_{kl}, \quad E\{\omega_k \nu_k^T\} = 0,
\]

where \( \delta_{kl} \) is the Kronecker delta function, and \( Q_k > 0, R_k > 0 \). The initial state \( x_0 \sim N(\tilde{x}_0, P_0) \) is independent of noises.

We consider the VB treatment of system (15) with an unknown input in both the state and measurement equations. The density of the measurement at time \( k \) is

\[
p(y_k | x_k, r_k) = N(y_k; H_k x_k + G_k r_k, R_k),
\]

and the complete-data likelihood is Gaussian.

Usually, the dynamic model of \( r_k \) is not known in detail, and thus some heuristic prior parameters can be employed for the dynamic evolution. Assume \( r_k \) has prior density \( N(r_k; \mu_k^-, \Sigma_k^-) \) by conjugacy. Then the VB learning results at time \( k - 1 \) are taken as the initial values of iteration at time \( k \) for the heuristic dynamic evolution of parameters. The structure of the considered system is illustrated in Fig. 1.

Using the known posterior density at time \( k - 1 \) and the convolution formula, we have the prediction densities:

\[
p(x_k | y_{1:k-1}) = N(x_k; \hat{\mu}_{k|k-1}, P_{k|k-1}),
\]
where  
\[ \hat{x}_{k|k-1} = A_{k-1} \hat{x}_{k-1|k-1} + F_{k-1} \mu_{k-1}, \]
\[ P_{k|k-1} = A_{k-1} P_{k-1|k-1} A_{k-1}^T + Q_{k-1}. \]

Taking expectation with respect to \( Q_x(x_k) \) on the following density
\[ p(y_k, x_k | r_k, y_{1:k-1}) \]
\[ \propto \exp \left( -\frac{1}{2} (x_k - \hat{x}_{k|k-1})^T P_{k|k-1}^{-1} (x_k - \hat{x}_{k|k-1}) \right. \]
\[ - \frac{1}{2} (y_k - H_k x_k - G_k r_k)^T R_k^{-1} (y_k - H_k x_k - G_k r_k), \]
\[ (16) \]
and applying (12), we have
\[ \ln Q_x(r_k) = \ln p(x_k, y_k | r_k, y_{1:k-1}) + \ln p(r_k) + \text{const} \]
\[ = -\frac{1}{2} r_k^T \left[ (\Sigma_k^{-1})^{-1} + G_k^T R_k^{-1} G_k \right] r_k \]
\[ + \left( \Sigma_k^{-1} \mu_k - G_k^T R_k^{-1} y_k \right)^T \left( \Sigma_k^{-1} \mu_k - G_k^T R_k^{-1} H_k x_k \right) x_k + \text{const}. \]

Let \( t_k = [r_k, y_k]$. Then \( p(t_k) \) is a multivariate Gaussian distribution. Thus we have the VB marginal distribution \( Q_v(r_k) = N(r_k; \mu_k, \Sigma_k) \) with
\[ \Sigma_k^{-1} = (\Sigma_k^{-1})^{-1} + G_k^T R_k^{-1} G_k, \]
\[ \Sigma_k^{-1} \mu_k = (\Sigma_k^{-1} \mu_k - G_k^T R_k^{-1} H_k x_k) x_k + \text{const}. \]

Similarly, from (11), we can obtain the VB marginal distribution \( Q_x(x_k) = N(x_k; \hat{x}_{k|k}, P_{k|k}) \) with
\[ P_{k|k}^{-1} = P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k, \]
\[ P_{k|k}^{-1} \hat{x}_{k|k} = P_{k|k-1}^{-1} \hat{x}_{k|k-1} + H_k^T R_k^{-1} (y_k - G_k \mu_k). \]

Under this circumstance, the optimal solution can be found analytically. In fact, equations (17) and (18) are coupled, and can be solved equivalently from the following equation:
\[ \left( \begin{array}{c} H_k^T \\ G_k^T \end{array} \right) R_k^{-1} \left[ \begin{array}{cc} H_k & G_k \\ 0 & \Sigma_k \end{array} \right]^{-1} \left[ \begin{array}{c} \hat{x}_{k|k} \\ \mu_k \end{array} \right] \]
\[ = \left[ P_{k|k-1}^{-1} \hat{x}_{k|k-1} + H_k^T R_k^{-1} y_k \right] \left( \Sigma_k^{-1} \mu_k - G_k^T R_k^{-1} H_k x_k \right). \]

We can see that, for the linear dynamic system (15), formulas (17) and (18) are similar due to the same density forms of \( x_k \) and \( r_k \). Specifically, if \( F_k = 0 \) and \( G_k = 0 \), then the proposed VB filter reduces to the standard Kalman filter.

**Remark 4.1:** Compared with the recursive three-step filter (RTSF) [15], the main difference lies in the way of estimating the input. The RTSF estimates the input by using the weighted least squares method, which treats errors between observation \( y_k \) and prediction \( H_k \hat{x}_{k|k-1} \) as the term of input. It may perform well when the measurement noise is small. However, when the measurement errors become largely caused by the measurement noise, the performance of RTSF would get worse quickly. On the other hand, the VB-filter estimates the state and input simultaneously.

**Remark 4.2:** The conjugate prior to (16) is the distribution of \( \eta_k \equiv R_k^{-1} G_k r_k \). If it is an equivalent transformation from \( \eta_k \) to \( r_k \), (17) and (18) can be derived directly by equations (13) and (14).

V. APPLICATIONS

This section addresses application of the proposed VB filtering algorithm to some practical dynamic systems with unknown input.

A. Linear Gaussian System

We consider a linear dynamic system described in [17]. The system parameters of (15) are
\[ A_k = \begin{bmatrix} 0.9 & -0.1 & -0.4 \\ 0.1 & 0.5 & -0.1 \\ 0 & 0.8 & 0 \end{bmatrix}, \]
\[ H_k = I_{3 \times 3}, \]
\[ F_k = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \]
\[ G_k = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \]

The noise covariance matrices are \( Q_k = \text{diag}(1, 1, 1) \) and \( R_k = \text{diag}(1, 0.1^2, 1) \). The initial state is \( x_0 \sim N(\tilde{x}_0, P_0) \), where \( \tilde{x}_0 = [0, 0, 0]^T \) and \( P_0 = I_{3 \times 3} \). The unknown input is partly modified by
\[ r_k = \begin{bmatrix} 4 u_s(k) - 4 u_s(k - 30) + 4 u_s(k - 65) \\ 3 u_s(k) - 3 u_s(k - 45) + 3 u_s(k - 75) \end{bmatrix}, \]
where \( u_s(\cdot) \) is the unit-step function. Due to the lack of information about the input, we assume a rather diffuse prior as \( r_1 \sim N(\mu_1, \Sigma_1^2) \), where \( \mu_1 = [0, 0]^T \) and \( \Sigma_1^2 = \text{diag}(4^2, 4^2) \).

The total simulation time was 100 steps. The tracking results of root mean square error (RMSE) using both the proposed VB-filter and the extended recursive three-step filtering (ERTSF) [17] are illustrated in Fig. 2. Table I lists the time-averaged RMSE of the state \( x = [x^{(1)}, x^{(2)}, x^{(3)}]^T \) and the trace of state estimates error covariance matrix \( P \) from 100 Monte Carlo simulations.

| Table I | Performance of State Estimation |
|---|---|---|---|---|
| Covariance Method | RMSE | uP |
| | \( x^{(1)} \) | \( x^{(2)} \) | \( x^{(3)} \) |
| \( 0.5 R \) | VB-filter | 0.596 | 1.292 | 0.614 | 2.408 |
| | ERTSF | 0.596 | 1.661 | 0.696 | 3.617 |
| \( R \) | VB-filter | 0.769 | 1.298 | 0.779 | 2.892 |
| | ERTSF | 0.771 | 1.831 | 0.982 | 4.937 |
| \( 2 R \) | VB-filter | 0.996 | 1.315 | 0.959 | 3.661 |
| | ERTSF | 1.005 | 2.172 | 1.437 | 7.836 |

We can see that both methods give good estimates which are close to the real state on average, but the VB-filter clearly has small errors in each component of the state. Note that only components \( x^{(2)} \) and \( x^{(3)} \) of the state are disturbed by the unknown input in this dynamic system. When the covariance of the measurement noise increases from \( 0.5 R \) to \( 2 R \), the RMSEs of undisturbed component \( x^{(1)} \) by the two methods
where \( \sigma \) increases rapidly for distributed components \( x^{(2)} \) and \( x^{(3)} \).

**B. Nonlinear Gaussian State Space Model**

Consider a so-called univariate non-stationary growth model, which has been used as a benchmark in [20]. The nonlinear dynamic model is slightly modified by adding the model, which has been used as a benchmark in [20]. During time

\[
x_k = \alpha x_{k-1} + \beta \frac{x_{k-1}}{1 + x_{k-1}^2} + \gamma \cos(1.2(k-1)) + F_k r_k + \omega_k,
\]

where \( \omega_k \sim N(0, \sigma^2_\omega), \nu_k \sim N(0, \sigma^2_\nu) \), and \( \omega_k \) and \( \nu_k \) are independent. In this simulation, we set \( \sigma^2_\omega = 1, \sigma^2_\nu = 1, \alpha = 0.5, \beta = 25, \gamma = 8, F_k = 0.5 \), and \( G_k = 2 \). The unknown input is \( r_k = 4u_k(k) - 4u_k(k-30) + 4u_k(k-60) \).

Since the complete-data likelihood of the nonlinear dynamic system is also Gaussian, we impose a rather diffuse Gaussian prior on the unknown input as \( r_k \sim N(0, 4^2) \) by conjugacy.

The total simulation time was 100 steps with initial state \( x_0 \sim N(0.1, 1^2) \). Fig. 3 shows the RMSE of the state estimates by the proposed VB-filter and the extended Kalman filter (EKF). EKF completely ignores the uncertainties. Overall, the proposed method obviously performs well because uncertainties were taken into consideration. During time 30–60, both methods give close state estimation results, since the unknown input decreased to zero in this period.

Table II lists the time-averaged RMSE for the proposed method and the EKF from 50 Monte Carlo simulations. The advantages of the proposed method in state estimates are clear.

**C. Tracking with Mixture Gaussian Noise**

Consider an object moving in a two-dimensional area tracked by a sensor. The dynamic system of the object is

\[
x_{k+1} = A_k x_k + \omega_k,
\]

\[
y_k = H_k x_k + \nu_k,
\]

where \( x_k \in R^n, y_k \in R^p \). Suppose \( \omega_k \) and \( \nu_k \) are independent, and \( \omega_k \sim N(0, Q_k), \nu_k \sim \sum_{m=1}^M \xi_m N(\mu^{(m)}_k, R^{(m)}_k) \), where \( \nu_k \) has a mixture Gaussian distribution with a fixed \( M \) components but unknown mixture coefficients \( \xi = [\xi_1, \ldots, \xi_M]^T \).

Consider the latent variable \( r_k = [r^{(1)}_k, \ldots, r^{(M)}_k]^T \), where \( r^{(m)}_k \in \{0, 1\} \) and \( \sum_{m=1}^M r^{(m)}_k = 1 \). Given mixture coefficients \( \xi \), the conditional density of \( r_k \) is

\[
p(r_k | \xi) = \prod_{m=1}^M \xi^r^{(m)}_m.
\]

Then the observation distribution becomes

\[
p(y_k | x_k, r_k) = \prod_{m=1}^M N(y_k | H_k x_k, R^{(m)}_k)^{r^{(m)}_k},
\]

and we have

\[
p(y_k | x_k, \xi) = p(y_k | x_k, r_k) p(r_k | \xi) = \sum_{m=1}^M p(y_k | x_k, \xi_m) p(\xi_m).
\]

And the likelihood is

\[
p(y_k, x_k | r_k) \propto \exp \left( - \frac{1}{2} \left( x_k - \hat{x}_{k|k-1} \right)^T P^{-1}_{k|k-1} \left( x_k - \hat{x}_{k|k-1} \right) \right).
\]

where

\[
V = \begin{bmatrix}
(y_k - H_k x_k)^T R^{(1)}_k (y_k - H_k x_k) \\
\vdots \\
(y_k - H_k x_k)^T R^{(M)}_k (y_k - H_k x_k)
\end{bmatrix}.
\]

Note that the Gaussian mixture distribution does not belong to the exponential family, but after importing random vectors

**TABLE II**

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE of ( x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB-filter</td>
<td>6.622</td>
</tr>
<tr>
<td>EKF</td>
<td>20.315</td>
</tr>
</tbody>
</table>
The complete-data likelihood $p(y_k,x_k|r_k)$ has the form of (7). Assuming factorization $Q(x_k,r_k,\xi) \approx Q_x(x_k)Q(r_k,\xi)$, the procedure of (13) and (14) at each time $k$ becomes:

1) Optimization of $Q_x(x_k)$ for fixed $Q(r_k,\xi)$. We use the standard Kalman filter to estimate the posterior distribution $Q^r_x(x_k)$ with measurement noise $\langle \xi, x_k \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the expectation with respect to $Q(r_k,\xi)$.

2) Optimization of $Q(r_k,\xi)$ for fixed $Q_x(x_k)$. We use the VB method again by factorization $Q(r_k,\xi) \approx Q_r(r_k)Q_\xi(\xi)$. The conjugate prior of $p(r_k)$ follows the Dirichlet distribution, and $p(\xi)$ follows the Dirichlet distribution $p(\xi;\alpha_0) = \frac{1}{B(\alpha_0)} \prod_{m=1}^{M} \xi^{\alpha_0(m)-1}$, where $B(\alpha_0) = \prod_{m=1}^{M} \Gamma(\alpha_0(m)) / \Gamma(\sum_{m=1}^{M} \alpha_0(m))$ is the normalization constant. Given observation $y_k$, we have the marginal posterior $Q_r(r_k) \propto \prod_{m=1}^{M} \left( \eta_k^{(m)}(r_k) \right)^{r_k(m)}$, $Q_\xi(\xi) = p(\xi;\alpha_k)$,

$$
\ln \eta_k^{(m)} = \langle \ln \xi \rangle_{r_k} - \frac{1}{2} \ln |R_k^{(m)}| - \frac{n}{2} \ln (2\pi)
$$

$$
- \frac{1}{2} (y_k - H_k \langle x_k \rangle_{x_k})^T R_k^{(m)^{-1}} (y_k - H_k \langle x_k \rangle_{x_k}),
$$

$$
\alpha_k = \alpha_{k-1} + \tau_k, \quad \tau_k = [\tau_1^{(1)}, \ldots, \tau_M^{(M)}]^T,
$$

$$
\bar{r}_k^{(m)} = \frac{\eta_k^{(m)}}{\sum_{m=1}^{M} \eta_k^{(m)}}, \quad m = 1, \ldots, M.
$$

We set $M = 3$, and the system parameters used are

$A_k = \begin{bmatrix} \cos(2\pi/300) & \sin(2\pi/300) \\ -\sin(2\pi/300) & \cos(2\pi/300) \end{bmatrix}$,

$H_k = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, $Q_k = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$,

$\mu_k^{(1)} = [1, 5]^T$, $\mu_k^{(2)} = [5, 1]^T$, $\mu_k^{(3)} = [8, 8]^T$,

$R_k^{(1)} = \begin{bmatrix} 100 & 0 \\ 0 & 20 \end{bmatrix}$, $R_k^{(2)} = \begin{bmatrix} 20 & 0 \\ 0 & 100 \end{bmatrix}$, $R_k^{(3)} = \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}$,

$\xi = [0.25, 0.35, 0.4]^T$, $\alpha_0 = [1/3, 1/3, 1/3]^T$, $\bar{x}_0 = [50, 0]^T$, $P_0 = \text{diag}(1^2, 1^2)$.

The following methods were performed for comparison:

- IMM filter: The interacting multiple model filter with transition probability matrix $[\xi, \xi, \xi]^T$;
- KFI: The Kalman filter with observation noise $v_k \sim N(0, Q_k)$.

Running 300 steps, the RMSE of the state using the VB-filter and IMM filters are illustrated in Fig. 4. The two methods have similar estimates close to the real positions of object.

To make further comparison, Table III lists the time-averaged RMSE of state and the trace of the state estimates error covariance matrix $P$ with 50 Monte Carlo simulations.

The IMM filter gives more accurate estimation for object’s position than other methods. Among the other methods without transition probabilities, the VB-filter has the lowest estimation errors. The last three Kalman filters perform poorly due to uncertainties ignored.

![Fig. 4. The RMSE of State Estimates](image-url)

<table>
<thead>
<tr>
<th>Filter</th>
<th>RMSE of $x^{(1)}$</th>
<th>RMSE of $x^{(2)}$</th>
<th>$\tau P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB-filter</td>
<td>5.060</td>
<td>4.719</td>
<td>48.029</td>
</tr>
<tr>
<td>IMM filter</td>
<td>4.896</td>
<td>4.578</td>
<td>45.030</td>
</tr>
<tr>
<td>KFI</td>
<td>5.496</td>
<td>5.238</td>
<td>57.778</td>
</tr>
<tr>
<td>KF2</td>
<td>5.435</td>
<td>5.121</td>
<td>55.876</td>
</tr>
<tr>
<td>KF3</td>
<td>5.705</td>
<td>4.604</td>
<td>53.869</td>
</tr>
</tbody>
</table>

**D. Tracking with Cluttered Measurements**

Consider a tracking system of a single target in two dimensions with cluttered measurements [16]. Suppose the dynamic model of the object is

$$x_{k+1} = A_k x_k + \omega_k, \quad \omega_k \sim N(0, Q_k),$$

where the state of the target is $x_k = [p_k^x, p_k^y, v_k^x, v_k^y]^T \in \mathbb{R}^4$, $(p_k^x, p_k^y)$ and $(v_k^x, v_k^y)$ are the position and velocity vectors in two dimensional Cartesian coordinates, respectively.

The likelihood of clutter measurements is assumed uniformly distributed in area $[-5, 5] \times [-4, 4]$. The joint likelihood of the system measurement is

$$p(y_k|x_k, \xi) = \begin{cases} 1/80, & \text{if } r_k = 0; \\ N(y_k; H_k \langle x_k \rangle_{x_k}, R_k), & \text{with probability } 1 - \xi. \end{cases}$$

We define the data association indicator $r_k$ at time $k$, where $r_k = 0$ if the measurement is clutter, and $r_k = 1$ if the measurement is from the actual target at time $k$. The prior of $r_k$ is assumed independent of the previous data associations and can be described as

$$p(r_k) = \begin{cases} \xi, & \text{if } r_k = 0; \\ 1 - \xi, & \text{if } r_k = 1. \end{cases}$$

For the unknown probability of $\xi$, we can use the VB filtering. By conjugacy, the prior of data association indicator
is assumed a multinomial distribution, and the prior of clutter is assumed a Dirichlet distribution.

The system parameters used are

\[
A_k = \begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad H_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},
\]

\[
Q_k = \begin{bmatrix} \frac{1}{2} & \Delta t^2 & 0 \\ 0 & \frac{1}{2} \Delta t & 0 \\ 0 & 0 & \Delta t \end{bmatrix}, \quad R_k = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}, \quad \alpha_0 = \frac{1}{2}.
\]

\[\dot{x}_0 = [-4, -0.2, 1, 0]^T, \quad P_0 = \text{diag}(0.1, 0.1, 0.1, 0.1),\]

where the step size was \(\Delta t = 0.1\), and the parameter in process noise was \(q = 0.1\). In this simulation, the probability of clutter was set as \(\xi = 0.35\). Running 240 steps, the real trajectory, state estimates performed by the VB-filter and the Rao-Blackwell particle filter (RBPF) [16], and the (object and clutter) measurements are plotted in Fig. 5.

![Fig. 5. Tracking Results with Cluttered Measurements](image)

To make comparison of these methods, Table IV lists the RMSE and trace of the estimates error covariance matrix from 50 Monte Carlo simulations. The two filters have comparable state estimation performance, because although no information about clutter possibility is given to the proposed method, the VB algorithm can give a close estimate of \(\xi\).

### E. Sensor Registration

In a multi-sensor tracking system, sensor registration errors, such as biased measurements, would possibly dramatically degrade estimation performance [31].

Consider a tracking scenario where a moving target is being tracked by \(l\) sensors [30]. The dynamic system concludes

<table>
<thead>
<tr>
<th>Method</th>
<th>(p^x)</th>
<th>(v^x)</th>
<th>(p^y)</th>
<th>(v^y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB-filter</td>
<td>3.1584</td>
<td>0.0030</td>
<td>2.0953</td>
<td>0.0074</td>
</tr>
<tr>
<td>KF</td>
<td>5.8113</td>
<td>0.0068</td>
<td>4.4092</td>
<td>0.0197</td>
</tr>
</tbody>
</table>

According to (17) and (18), we can perform filtering methods by augmenting system parameters. In this simulation, Sensor 1 is located at \((0, 0)\) with sensor registration bias \(\hat{\rho}^{(1)} = 10, \hat{\theta}^{(1)} = 0.0105\), and Sensor 2 is located at \((300, 0)\) with bias \(\hat{\rho}^{(2)} = -8, \hat{\theta}^{(2)} = 0.0087\).

The system was simulated 200 steps with \(\Delta t = 3, \sigma^2_{v_x} = 0.00035, \sigma^2_{v_y} = 0.00035\). The initial state was given as \(x_0 = N(\bar{x}_0, P_0)\), where \(\bar{x}_0 = [230, -0.2, 500, 0.1]^T\) and \(P_0 = \text{diag}(10^2, 0.1^2, 10^2, 0.1^2)\). For comparison, the Kalman filtering fusion algorithm was also performed, which ignored the sensors’ biases. The tracking results are shown in Fig. 6.

Table V gives the time-averaged RMSE from 100 Monte Carlo simulations. It shows the improvement by handling measurement bias.

<table>
<thead>
<tr>
<th>Method</th>
<th>(p^x)</th>
<th>(v^x)</th>
<th>(p^y)</th>
<th>(v^y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB-filter</td>
<td>3.1584</td>
<td>0.0030</td>
<td>2.0953</td>
<td>0.0074</td>
</tr>
<tr>
<td>KF</td>
<td>5.8113</td>
<td>0.0068</td>
<td>4.4092</td>
<td>0.0197</td>
</tr>
</tbody>
</table>

989
proposed methods. and related numerical simulations show the advantages of the input by a closed-form recursive procedure. Some applications variational Bayesian filter can estimate the system state and gate priors are incorporated to quantify the uncertainties. The input. With no prior dynamic properties of the input, the conjugu-

[3] Y. Bar-Shalom and X. R. Li, Applications and Advances
Blackwellized data association particle filters,” Documentation of RBM-
[12] X. Gao, J. Chen, D. Tao, and X. Li, “Multi-sensor centralized fusion without measurement noise covariance by variational Bayesian approxi-

VI. CONCLUSION
This paper presents a recursive variational Bayesian filter to estimate the state of a linear dynamic system with unknown input. With no prior dynamic properties of the input, the conjugate priors are incorporated to quantify the uncertainties. The variational Bayesian filter can estimate the system state and input by a closed-form recursive procedure. Some applications and related numerical simulations show the advantages of the proposed methods.

REFERENCES
[12] X. Gao, J. Chen, D. Tao, and X. Li, “Multi-sensor centralized fusion without measurement noise covariance by variational Bayesian approxima-

Fig. 6. Biased Measurements and Fusion Results