Variational Compensation Based Nonlinear Filter for Continuous-Discrete Stochastic Systems

Tingjun Wang, Haoran Cui and Xiaoxu Wang

July 6, 2020

Variational Compensation Based Nonlinear Filter for Continuous-Discrete Stochastic Systems

1st Tingjun Wang
Institute No.16 of China Aerospace science and technology corporation
Xi'an, Shaanxi province, China
wangtingjun0309@126.com

2st Haoran Cui
Northwestern Polytechnical University School of Automation
Xi'an, China
the_last_cowboy@163.com

3nd Xiaoxu Wang
Northwestern Polytechnical University School of Automation
Xi'an, China
woyaofly1982@163.com

***Abstract*—In this paper, a novel variational compensation based nonlinear filter (VCNF) is proposed to cope with the nonlinear filtering problem in continuous-discrete systems. The core of VCNF is to construct a variational state compensation model with variational compensation parameters for accurately describing uncertain continuous state. The role of variational compensation parameters is to adaptively compensate the unpredictable approximation and discretization errors of system states. In the variational Bayesian framework, through iteratively and alternatively achieving the fitting of the state priori model and the compensation of approximation and discretization errors, estimation accuracy and adaptiveness can be enhanced gradually. The superior performance of VCNF is demonstrated in the simulation of target tracking.**

***Index Terms*—Variational Bayesian method, Continuous-discrete stochastic system, Nonlinear Kalman filter, Target tracking**

I. Introduction

In continuous-discrete system, process dynamics are governed by a continuous-time stochastic differential equation (SDE) and measurements are taken at discrete time instants, which can accurately reflect the real engineering problems. The filtering problems of continuous-discrete systems also widely exist in target tracking and navigation [1], [2], stochastic control [3], fault estimation [4], systems biology [5] and many other real-life estimation problems.

For filtering problem of linear continuous-discrete systems, continuous-discrete-Kalman filter (CD-KF) is the optimal solution [6], where the dynamic state is described by an SDE and the measurements are at discrete times. In continuous-discrete nonlinear problems, many variants of CD-KF are proposed, which share the same framework, i.e. the time update and measurement update. The authors of [7] discussed two different kinds of numerical approximation methods in CD-EKF, i.e. numerical integration and discrete approximation, for solving the moment differential equation (MDE) in time update. But either in integration or in discretization, the step size is fixed, which will make the CD-EKF procedure quickly diverge and limit the application so that many advanced CD-EKF with variable-step size are proposed [8], [9]. Besides CD-EKF, various implementations of CD-UKF [10] and CD-CKF [11]–[13] were discussed as well and classified into the discretization and integration two categories. However, the arising unpredictable numerical errors may broke the positive definiteness of covariance and lead to degradation of filtering performance [14]. Hence, in the accurate CD-EKF (ACD-EKF) [15], an efficient embedded Runge-Kutta pair was applied to process global error and simplify computational procedure. Especially for the radar tracking, the ACD-EKF is compared with CD-UKF and CD-CKF [16] to demonstrate the impact of error control. Accurate CD-CKF (ACD-CKF) was proposed in [17], but the involved iteration processes for integrating matrix differential equation are quite time-consuming. In [18] an advance ACD-CKF is designed which is particularly effective for CD stochastic systems with nonlinear and/or non-differentiable observations. When the continuous-discrete system is not Gaussian system, particle filter can be referred and extended to cope with it and develop CD-particle filter (CD-PF) [19], [20]. However, in the existing filters for nonlinear continuous-discrete systems, the discretization of differential equation and approximation of nonlinear integral are both inevitable. The unpredictable error and uncertainty caused by approximation and discretization will greatly influence the accuracy and confidence of estimation results. Especially for rare measurement with long interval of measurement, state have to suffer from long-term propagation without measurement to correct state. An inaccurate and unreliable state mean and covariance even can lead to divergency. Moreover, for decreasing the discretization error, the discretization interval of state must be decreased as well so that the computational burden will be increased.

This paper aims to overcome the above-mentioned disadvantages and a novel variational compensation nonlinear filter (VCNF) is proposed. The main challenge of VCNF is how to accurately model the continuous system state priori model and adaptively compensate the approximation and discretization error. In this paper, we construct variational state compensation model (VSCM) to model the state priori probability. In VSCM, variational compensation parameters (VCPs) are introduced in both mean and covariance of VSCM. Instead of single point, VCPs represent probability distributions, which can reflect and capture more information about uncertain continuous state. In VCNF, the role of VSCM is to accurately describe the uncertain state priori model and VCPs are used to adaptively compensate the approximation and discretization error and adjust the confidence of VSCM. Based on VSCM, the model of complete-data likelihood probability (CLP) can be established as well. Then, given the model of CLP, through iteratively maximizing evidence lower bound (ELBO) in variational bayesian (VB) framework, the identification of VCPs, the fitting of state priori model and the estimation of state can be achieved jointly. This paper's contributions lie in that: 1) by characterizing the state priori model with VSCM, the uncertainty of continuous system state can be reflected so that the adaptiveness and robustness of VCNF to approximation and discretization are strong. 2) By introducing a set of VCPs in the mean and covariance of VSCM, different errors can be compensated and the estimation accuracy can be improved gradually.

The rest of this paper is organized as follows: the problem is formulated and our idea is briefly introduced in section II. In section III, the models of VSCM and CLP are established and the iterative maximization process of ELBO is derived. Section IV gives the simulation results illustrating that our proposed VCNF is superior. Concluding remarks are given in section V.

II. Problem Formulation

In this paper, the nonlinear state-space model of the continuous-discrete stochastic system [6] is considered as

$$dx\left(t\right)=f\left(x\left(t\right),t\right)dt+Gdw\left(t\right),\quad t>0 \tag{1}$$

where $x(t) \in \mathbb{R}^{\mathrm{n}}$ represents the n-dimensional continuous state of the system at time t; $\{w(t), t>0\}$ denotes the n-dimension standard Brownian motion with $E\left[dw\left(t\right)dw(t)^{\mathrm{T}}\right] = Q\left(t\right)dt$ that is independent of $x(t)$; $G \in {}^{n\times n}$ is the gain matrix of the process noise and $f : \mathbb{R}\times\mathbb{R}^{\mathrm{n}} \rightarrow \mathbb{R}^{\mathrm{n}}$ is a known differentiable nonlinear drift function.

The behavior of the system is observed by noisy measurements at sampled time instants $t_k = k\Delta t$, i.e.

$$z_k = h(x_k) + v_k \tag{2}$$

$z_k \in \mathbb{R}^{\mathrm{m}}$ is the measurement at discrete instant $t = t_k$ with respect to $x_k = x(t_k)$. The discrete measurements are assumed to arrive with equidistant interval between z_k and z_{k-1}, i.e. $\Delta t = t_k - t_{k-1}$. $h : \mathbb{R}^{\mathrm{n}}\times\mathbb{R} \rightarrow \mathbb{R}^{\mathrm{m}}$ is nonlinear differentiable measurement function. Measurement noise v_k is zero-mean Gaussian white noises satisfying $\mathrm{E}[v_k v_j^{\mathrm{T}}] = \delta_{kj}R$ and $R > 0$. The initial state is assumed to obey Gaussian with mean $\hat{x}_0$ and covariance P_0, which is independent of $w(t)$ and v_k.

The aim of continuous-discrete filter is to accurately calculate the posterior distributions of current state $x(t)$ given all history of the measurements, i.e.

$$p\left(x\left(t\right)|z_1, z_2, \ldots, z_k\right), t\in\left[t_k, t_{k+1}\right), k = 1, 2, \ldots \tag{3}$$

Compared with traditional discrete filtering problem, the measurement update step of continuous-discrete filter is the same as that of relative discrete-discrete filter, which does not need to be discussed in this paper. The main difference of continuous-discrete filter is the probability of the continuous state $x(t) \sim N(\hat{x}(t), P(t))$, which is actually defined for all $t \geq 0$ and not only for the discrete measurement steps at t_k.

Hence, in continuous-discrete filtering, the propagations of system state $\hat{x}(t)$ and covariance $P(t)$ at state prediction step without arrived measurements require first solving the full Fokker-Planck-Kolmogorov (PFK) partial differential equations [4], [6], i.e.

$$\begin{aligned}
\frac{d\hat{x}\left(t\right)}{dt} =& E\left[f\left(x\left(t\right),t\right)\right] \qquad (4)\\
\frac{dP\left(t\right)}{dt} =& E\left[\left(x\left(t\right)-\hat{x}\left(t\right)\right)f^{\mathrm{T}}\left(x\left(t\right),t\right)\right]\\
&+ E\left[f\left(x\left(t\right),t\right)\left(x\left(t\right)-\hat{x}\left(t\right)\right)^{\mathrm{T}}\right] + GQ\left(t\right)G^{\mathrm{T}}\\
=& P\left(t\right)E[F_x\left(x\left(t\right),t\right)]^{\mathrm{T}}\\
&+ E[F_x\left(x\left(t\right),t\right)]^{\mathrm{T}}P\left(t\right) + GQ\left(t\right)G^{\mathrm{T}} \qquad (5)
\end{aligned}$$

where $t \in [t_k, t_{k+1}]$; $E[\cdot]$ represents the expectation with respect to $x(t) \sim N(\hat{x}(t), P(t))$. $F_x(x(t), t)$ is the Jacobian matrix of $f(x(t), t)$ with respect to $x(t)$, with elements $[F_x]_{ij} = \frac{\partial f_i}{\partial x_j}$.

A. Continuous-discrete nonlinear filter

Because of the nonlinear drift and measurement function, it is almost impossible to obtain analytical and closed-form solution of these equations (4)-(5) at state prediction step. Hence, in order to approximately implement the differential equations (4)-(5) on the state prediction step, the unpredictable error and uncertainty will be inevitably caused by two main reasons.

The first reason is the approximation of expectation in (4)-(5) with the general form, i.e.

$$E\left[g\left(x,t\right)\right] = \int g\left(x,t\right)N\left(x|m,P\right)dx \tag{6}$$

Two different kinds of popular approximation methods are Taylor expansion [21] and sigma-point (unscented transform and Cubature method) [10], [11] methods, which will lead to CD-EKF, CD-UKF and CD-CKF, respectively. Obviously other approximations exist as well. For example, it is also possible to use sequential Monte Carlo based particle filter [20] and Markov chain Monte Carlo based method [22]. However, in these approaches, the approximation error of the expectation can not be avoided well.

The second reason is the discretization of differential equations in (4)-(5). The error caused by discretization mainly depend on the interval of discretization, i.e. the smaller the interval is, the smaller discretization error is. However, correspondingly small interval will lead to heavy computational burden.

Besides above mentioned PFK partial differential equation based methods, there exists another way to solve continuous-discrete filtering problem, i.e. directly using stochastic numerical scheme to simulate or discrete the stochastic differential equation (1). For example, in [11] the 1.5 order strong Itô-Taylor scheme is used to discrete stochastic differential equation (1). Then, the result of the state prediction step can be obtained through approximately calculating the expectation with the form in (6). However, in different kinds of continuous-discrete nonlinear filters, the unpredictable error and uncertainty caused by approximation of expectation and discretization are both inevitable and not negligible. For nonlinear system and rare measurements with large Δt, the approximation and discretization errors will demolish the performance of filter or even lead to divergence.

B. Our idea

In order to decrease above discussed errors and obtain accurate filtering results, in this paper, we introduce VCPs to be iteratively optimized at the state prediction step for more accurately approximating the continuous state probability.

To this end, we construct VSCM with VCPs to describe the true state priori model, i.e.

$$p\left(x\left(t_{k,\sigma}\right)|x\left(t_{k,\sigma-1}\right)\right) \approx p\left(x\left(t_{k,\sigma}\right)|x\left(t_{k,\sigma-1}\right),\Lambda_k,\Sigma_k\right) \quad (7)$$

where $\sigma = 1, 2, \ldots, \theta$ and θ is the number of state propagation during a measurement interval Δt. $\alpha = \frac{\Delta t}{\theta}$ is the interval of discretization of continuous state. For notation concision, in this paper, $t_{k,\sigma}$ denotes $t_{k+\sigma\alpha}$. When $\sigma = \theta$, $x\left(t_{k-1,\sigma}\right) = x\left(t_k\right)$

In (7), Λ and Σ are VCPs to be optimized for enhancing adaptivity and accuracy of VSCM. This paper aims to propose a novel nonlinear filter for continuous-discrete system, in which VCPs and system state can be iteratively identified and estimated based on VB framework. Through optimizing VSCM by identifying VCPs, more accurate state estimation can be achieved. Correspondingly, state estimation will also contribute to the identification of VCPs. The construction of VSCM and the derivation of VCNF will be shown in the following section.

III. Variational Compensation Nonlinear Filter

The first challenge of CD-VCNF is to design VSCM and properly introduce VCPs to compensate approximation and discretization error. Then, based on VB framework and proposed VSCM, the model optimization and state estimation can be achieved iteratively by maximizing ELBO.

A. Construction of the variational state compensation model

In order to more accurately describe the continuous state and control the negative influence caused by approximation and discretization error, we propose parametric VSCM to approximate the state priori model, i.e.

$$\begin{aligned}
&p\left(x\left(t_{k,\sigma}\right)|x\left(t_{k,\sigma-1}\right)\right) \approx \\
&N\left(x\left(t_{k,\sigma}\right)|F_v\left(\hat{x}\left(t_{k,\sigma-1}\right)\right)+\Lambda_k,\left(P_v\left(P\left(t_{k,\sigma-1}\right)\right)\Sigma_k\right)^{-1}\right) \\
&= p\left(x\left(t_{k,\sigma}\right)|\hat{x}\left(t_{k,\sigma-1}\right),\Lambda_k,\Sigma_k\right), \sigma = 1, 2, \ldots, \theta \qquad (8)
\end{aligned}$$

where

$$F_v\left(\hat{x}\left(t_{k,\sigma-1}\right)\right) = \hat{x}\left(t_{k,\sigma-1}\right) + \alpha f\left(\hat{x}\left(t_{k,\sigma-1}\right), t\right) \quad (9)$$

$$\begin{aligned}
&P_v\left(P\left(t_{k,\sigma-1}\right)\right)^{-1} \\
&= P\left(t_{k,\sigma-1}\right) + \alpha P\left(t_{k,\sigma-1}\right) F_x\left(\hat{x}\left(t_{k,\sigma-1}\right), t\right)^{\mathrm{T}} \\
&+ \alpha F_x\left(\hat{x}\left(t_{k,\sigma-1}\right), t\right) P\left(t_{k,\sigma-1}\right)^{\mathrm{T}} + \alpha G Q(t) G^{\mathrm{T}} \qquad (10)
\end{aligned}$$

Comparing (9)-(10) with (4)-(5), we can find that they are similar. Actually, using Euler method to discrete the differential equations (4)-(5) and Taylor expansion to approximate the expectation with the form of (6), we can obtain (9)-(10), which is employed as parts of mean and covariance of VSCM in (8).

However, besides $F_v\left(\hat{x}\left(t_{k,\sigma-1}\right)\right)$ and $P_v\left(P\left(t_{k,\sigma-1}\right)\right)$, the more important parts in are VCPs Λ and Σ. It is remarkable that VCPs Λ and Σ represent probability distributions, rather than single point. Because distribution can reflect more information than point, through optimizing VCPs, error and uncertainty of state can be captured and compensated more effectively by probability distributions. This is the main reason why we use VB framework, rather than other point-estimation methods. The role of VCP Λ is to compensate the error of state's mean. Σ is used to adjust the confidence of mean and reflect the fluctuation of error. Through iteratively optimizing Λ and Σ, the discretization and approximation error of state can be significantly reduced and the state priori model can be fitted well. Then, the accuracy of state estimation can be improved gradually.

Because VCPs are probability, to calculate the posteriori probability of VCPs, their priori probabilities need to be defined. The following analysis can be considerably simplified if conjugate prior distributions are selected. We therefore use Gaussian-Wishart prior governing the mean and precision of VSCM. i.e.

$$\begin{aligned}
p\left(\Lambda_k,\Sigma_k\right) &= p\left(\Lambda_k|\Sigma_k\right)p\left(\Sigma_k\right) \\
&= N\left(\Lambda_k|\Lambda_0,\left(\lambda_0\Sigma_k\right)^{-1}\right)\omega\left(\Sigma_k|W_{\Sigma_0},V_{\Sigma_0}\right) \quad (11)
\end{aligned}$$

The hyperparameters Λ_0^i, λ_0^i, W_{Σ_0}, V_{Σ_0} need to be initialized at the beginning of our proposed algorithm.

As for the process of measurement function, we will use the Taylor expansion to approximate the measurement likelihood probability $p\left(z_k|x\left(t_k\right)\right)$, i.e.

$$p\left(z_k|x\left(t_k\right)\right) \approx N\left(z_k|H_k x\left(t_k\right)+u_k,R_k\right) \qquad (12)$$

where $H_k \in \mathrm{R}^{m\times n}$ and $u_k \in \mathrm{R}^m$ are the priori known matrix and vector respectively, which can be considered as Hessian matrix and first-order constant term of the Taylor expansion, i.e.

$$\begin{aligned}
H_k &= \frac{\partial^2 h\left(x_k\right)}{\partial {x_k}^2} \\
u_k &= h\left(\hat{x}_{k/k-1}\right) - H_k\hat{x}_{k/k-1}
\end{aligned}$$

where $\hat{x}_{k/k-1}$ is the result of state prediction step at sampling time t_k.

B. Variational iterative calculation of posteriori probability

Employing VB approach, posteriori probability can be calculated by maximizing the ELBO [23], i.e.

$$L(q) = \int q\left(x\left(t_k\right), \Lambda_k, \Sigma_k\right) \times \ln \frac{p\left(Z_1^k, x\left(t_k\right), \Lambda_k, \Sigma_k\right)}{q\left(x\left(t_k\right), \Lambda_k, \Sigma_k\right)} d\left\{x\left(t_k\right), \Lambda_k, \Sigma_k\right\} \quad (13)$$

where $q\left(x_k, \Lambda_k, \Sigma_k\right)$ denotes posteriori probability of state and VPs, which can be factorized with respect to these groups, i.e.

$$q\left(x\left(t_k\right), \Lambda_k, \Sigma_k\right) = q\left(x\left(t_k\right)\right) q\left(\Lambda_k|\Sigma_k\right) q\left(\Sigma_k\right) \quad (14)$$

Based on VSCM in (8), measurement likelihood probability in (12) and priori probabilities of VCPs in (11), the complete-data likelihood probability $p\left(Z_1^k, x\left(t_k\right), \Lambda_k, \Sigma_k\right)$ in (13) can be expressed as

$$\begin{aligned} &p\left(Z_1^k, x\left(t_k\right), \Lambda_k, \Sigma_k\right) \\ &= p\left(z_k|x\left(t_k\right)\right) p\left(x\left(t_k\right)|\hat{x}\left(t_{k-1}\right), \Lambda_k, \Sigma_k\right) \\ &\times p\left(\Lambda_k|\Sigma_k\right) p\left(\Sigma_k\right) p\left(Z_1^{k-1}\right) \end{aligned} \quad (15)$$

Based on the model of complete-data likelihood probability in (15), the posterior probabilities of the state $q\left(x\left(t_k\right)\right)$ and VCPs $q\left(\Lambda|\Sigma\right) q\left(\Sigma\right)$ can be iteratively calculated by maximizing the ELBO $L\left(q\right)$ in (13) with model identification stage and state correctness stage, which are derived as follows.

1) model identification stage: Given the posterior probability of state $x\left(t_k\right)$ in the i-th iteration, i.e.

$$q_i\left(x\left(t_k\right)\right) = N\left(x\left(t_k\right)|\hat{x}_i\left(t_k\right), P_i(t_k)^{-1}\right)$$

then, we can calculate the posterior probabilities of VCPs Λ, Σ in the i+1-th iteration, i.e.

$$\begin{aligned} q_{i+1}\left(\Lambda_k|\Sigma_k\right) &= N\left(\Lambda_k|\Lambda_k^{i+1}, \left(\lambda_k^{i+1}\Sigma_k\right)^{-1}\right) \\ q_{i+1}\left(\Sigma_k\right) &= \omega\left(\Sigma_k|W_{\Sigma_k}^{i+1}, V_{\Sigma_k}^{i+1}\right) \end{aligned}$$

where

$$\lambda_k^{i+1} = P_v\left(P\left(t_{k-1}\right)\right) + \lambda_0 \quad (16)$$

$$\Lambda_k^{i+1} = \left(\lambda_k^{i+1}\right)^{-1} P_v\left(P\left(t_{k-1}\right)\right)\left(\hat{x}_i\left(t_k\right) - F_v\left(\hat{x}\left(t_{k-1}\right)\right)\right) + \left(\lambda_k^{i+1}\right)^{-1}\lambda_0\Lambda_0 \quad (17)$$

$$\left(W_{\Sigma_k}^{i+1}\right)^{-1} = P_i(t_k)^{-1} P_v\left(P\left(t_{k-1}\right)\right) + \left(W_{\Sigma_0}\right)^{-1} + \bar{X}\bar{X}^{\mathrm{T}}\left(P_v\left(P\left(t_{k-1}\right)\right)^{-1} + \lambda_0^{-1}\right)^{-1} \quad (18)$$

$$V_{\Sigma_k}^{i+1} = V_{\Sigma_0} + 1 \quad (19)$$

$$\bar{X} = \hat{x}_i\left(t_k\right) - F_v\left(\hat{x}\left(t_{k-1}\right)\right) - \Lambda_0$$

Proof: According to the mean field theory used in VB method, based on the decomposition of in (15), it is easy to get

$$\begin{aligned} &\ln q_{i+1}\left(\Lambda_k, \Sigma_k\right) \\ &= E_{q_i(x(t_k))}\left\{\ln p\left(Z_1^k, x\left(t_k\right), \Lambda, \Sigma\right)\right\} + const \\ &= E_{q_i(x(t_k))}\left\{\ln p\left(x\left(t_k\right)|\hat{x}\left(t_{k-1}\right), \Lambda_k, \Sigma_k\right)\right\} \\ &+ \ln p\left(\Lambda_k|\Sigma_k\right) + \ln p\left(\Sigma_k\right) + const \end{aligned} \quad (20)$$

Note that we are only interested in the functional dependence of the right-hand side on the VCPs. Hence, any terms that do not depend on VCPs can be absorbed into the additive normalization constant in (20). Then, the expectation in (20) can be evaluated i.e.

$$\begin{aligned} &E_{q_i(x(t_k))}\left[\ln p\left(x\left(t_k\right)|\hat{x}\left(t_{k-1}\right), \Lambda_k, \Sigma_k\right)\right] \\ &= \frac{1}{2}\ln\left|\Sigma_k\right| - \frac{1}{2}\bar{X}^{\mathrm{T}} P_v\left(P\left(t_{k-1}\right)\right)\Sigma_k\bar{X} \\ &- \frac{1}{2}Tr\left[P_i\left(t_k\right) P_v\left(P\left(t_{k-1}\right)\right)\Sigma_k\right] \end{aligned} \quad (21)$$

Then given (11) and (21), by re-organizing (20), we can obtain (16)-(19).

■

2) state estimation stage: Given posterior probabilities of VCPs Λ, Σ in the i+1-th iteration, i.e.

$$\begin{aligned} q_{i+1}\left(\Lambda_k|\Sigma_k\right) &= N\left(\Lambda_k|\Lambda_k^{i+1}, \left(\lambda_k^{i+1}\Sigma_k\right)^{-1}\right) \\ q_{i+1}\left(\Sigma_k\right) &= \omega\left(\Sigma_k|W_{\Sigma_k}^{i+1}, V_{\Sigma_k}^{i+1}\right) \end{aligned}$$

then, we can calculate the posterior probability of state $x\left(t_k\right)$ in the i+1-th iteration, i.e.

$$q_{i+1}\left(x\left(t_k\right)\right) = N\left(x\left(t_k\right)|\hat{x}_{i+1}\left(t_k\right), P_{i+1}(t_k)^{-1}\right)$$

where

$$P_{i+1}\left(t_k\right) = P_v\left(P\left(t_{k-1}\right)\right) W_{\Sigma_k}^{i+1} V_{\Sigma_k}^{i+1} + H_k^{\mathrm{T}} R_k^{-1} H_k \quad (22)$$

$$\hat{x}_{i+1}\left(t_k\right) = P_{i+1}(t_k)^{-1} H_k^{\mathrm{T}} R_k^{-1}\left(z_k - u_k\right) + P_{i+1}(t_k)^{-1} W_{\Sigma_k}^{i+1} V_{\Sigma_k}^{i+1}\left(F_v\left(\hat{x}\left(t_{k-1}\right)\right) + \Lambda_k^{i+1}\right) \quad (23)$$

Proof: According to the mean field theory used in VB approach, based on the decomposition in (15), it is easy to get

$$\begin{aligned} &\ln q_{i+1}\left(x\left(t_k\right)\right) \\ &= E_{q_{i+1}(\Lambda_k, \Sigma_k)}\left\{\ln p\left(Z_1^k, x\left(t_k\right), \Lambda, \Sigma\right)\right\} + const \\ &= E_{q_{i+1}(\Lambda_k, \Sigma_k)}\left\{\ln p\left(x\left(t_k\right)|\hat{x}\left(t_{k-1}\right), \Lambda_k, \Sigma_k\right)\right\} \\ &+ \ln p\left(z_k|x\left(t_k\right)\right) + const \end{aligned} \quad (24)$$

In (24), we only pay attention to the functional dependence on the system state $x\left(t_k\right)$. Hence, terms that do not depend on state $x\left(t_k\right)$ can be considered as the additive normalization constant in (24). Then, the expectation in (24) can be evaluated i.e.

$$\begin{aligned} &E_{q_{i+1}(\Lambda_k, \Sigma_k)}\left\{\ln p\left(x\left(t_k\right)|\hat{x}\left(t_{k-1}\right), \Lambda_k, \Sigma_k\right)\right\} \\ &= -\frac{1}{2}\bar{X}_k^{\mathrm{T}} P_v\left(P\left(t_{k-1}\right)\right) W_{\Sigma_k}^{i+1} V_{\Sigma_k}^{i+1}\bar{X}_k + const \end{aligned} \quad (25)$$

where $\bar{X}_k = x\left(t_k\right) - F_v\left(\hat{x}\left(t_{k-1}\right)\right) - \Lambda_k^{i+1}$

Then given (12) and (25), by re-organizing (24), we can obtain (22)-(23). ■

The maximization process of the lower bound has been derived, which includes model identification and state estimation stages. From the updated formulation in every stage, it is obviously that the solution of maximizing ELBO has an analytical expression form. Through iteratively and alternatively operating these two stages, the error compensation, model identification and state estimation can be achieved simultaneously.

IV. SIMULATION

In this section, our proposed CD-VCNF is compared with CD-EKF, CD-UKF and CD-CKF in the target tracking nonlinear model.

The aircrafts dynamics obeys an stochastic differential equation of the form (1), where the state vector $\mathrm{x}(t) := [x(t), \dot{x}(t), y(t), \dot{y}(t), z(t), \dot{z}(t), \varphi(t)]^{\mathrm{T}} \in \mathbb{R}^7$. $x(t)$, $y(t)$, $z(t)$ and $\dot{x}(t)$, $\dot{y}(t)$, $\dot{z}(t)$ are positions and velocities at time t and $\varphi(t)$ denotes turn rate. The drift function is $f(\mathrm{x}(t)) := [x(t), -\varphi(t)\dot{y}(t), \dot{y}(t), \varphi(t)\dot{x}(t), \dot{z}(t), 0, 0]^{\mathrm{T}} \in \mathbb{R}^7$. The driving noise term is also 7-D with all entries $\{\varphi_i(t), t > 0\}$, $i = 1, 2, \ldots, 7$ being mutually independent Brownian processes with zero mean and unit covariance. Its matrix G is constant, diagonal, and given by the formula $G := diag[0, \sigma_1, 0, \sigma_1, 0, \sigma_1, \sigma_2]^{\mathrm{T}}$ with $\sigma_1 = \sqrt{0.2}m/s$ and $\sigma_2 = 0.007^\circ s^{-1}$.

The observation model in the explored air traffic control scenario is nonlinear, discrete-time, i.e.

$$h(x_k) = \begin{bmatrix} \sqrt{x_k^2 + y_k^2 + z_k^2} \\ \tan^{-1}\left({y_k}/{x_k} \right) \\ \tan^{-1}\left({z_k}\Big/{\sqrt{x_k^2 + y_k^2}} \right) \end{bmatrix} + \upsilon_k$$

where x_k, y_k, z_k denote the aircrafts position at time t_k. The measurement is polluted by the white noise $\upsilon_k \sim N(0, R)$, where with the constant diagonal covariance matrix $R = diag\left[\sigma_r^2, \sigma_\theta^2, \sigma_\varphi^2\right]$ and $\sigma_r = 50m$, $\sigma_\theta = 0.1^\circ$, $\sigma_\varphi = 0.1^\circ$.

In this simulation, the simulation time K is $100s$. Measurement interval Δt is $1s$. In order to adequate test estimation accuracy and robustness to error of different filters, the discretization interval α is set as $0.5s$. Initial state is the 7-D Gaussian variable, with mean $\left[1000m, 0m/s, 2560m, 150m/s, 200m, 0m/s, 3^\circ s^{-1}\right]^{\mathrm{T}}$ and covariance $diag\,[0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01,]$. Furthermore, the hyperparameters in our proposed CD-VCNF are chosen as

$$\begin{aligned} V_{\Sigma_0} &= 100 \quad W_{\Sigma_0} = 0.01 I_{n\times n} \\ \Lambda &= 0_{n\times 1} \qquad \lambda = 0.01 I_{n\times n} \end{aligned}$$

where I is the unit matrix.

In order to evaluate the performance of different filters, we used the root mean square error (RMSE) over $N_{mc} = 100$ simulation runs:

$$RMSE_k(t_k) = \sqrt{\frac{1}{N_{mc}} \sum_{l=1}^{N_{mc}} \left(\bar{x}_k^l\right)^2 + \left(\bar{y}_k^l\right)^2 + \left(\bar{z}_k^l\right)^2}$$
$$k = 1, 2, \cdots, K$$

where k is the sampling time of measurement and

$$\begin{aligned} \bar{x}_k^l &= x^l(t_k) - \hat{x}^l(t_k) \\ \bar{y}_k^l &= y^l(t_k) - \hat{y}^l(t_k) \\ \bar{z}_k^l &= z^l(t_k) - \hat{z}^l(t_k) \end{aligned}$$

$x^l(t_k)$ is the true value of state and $\hat{x}^l(t_k)$ is the estimated value of state in l-th simulation at time t_k. Furthermore, the average RMSE (ARMSE) is considered as a evaluation index as well, i.e.

$$ARMSE[i] = \frac{1}{K} \sum_{k=1}^{k} RMSE_k[i]$$

We also considered the mean absolute error (MAE) over time t_k of the l-th simulation run, i.e.,

$$MAE(l) = \frac{1}{k} \sum_{k=1}^{K} \left(\left|\bar{x}_k^l\right| + \left|\bar{y}_k^l\right| + \left|\bar{z}_k^l\right| \right)$$

First of all, we compare the estimation accuracy of our proposed CD-VCNF and other filters. To this end, the ARMSEs of position and velocity are shown in table I and II, respectively. Obviously, the ARMSEs of position and velocity of CD-VCNF are both the lowest and the worst one is CD-UKF. To be more specific, the curves of RMSEs of position and velocity over 100s are described in Fig.1-2. We can find that not only the estimation accuracy of CD-VCNF is higher than that of other filters, but also the fluctuation of RMSE of CD-VCNF is also the least. Around 20 and 80 seconds in Fig.1 and around 90 second in Fig.2, there is upward trend in RMSEs of CD-UKF and CD-CKF. However, RMSEs of CD-VCNF is relative stable over 100s.

In order to further analyze the estimation stability in the 100 Monte Carlo simulations, the MAEs of position and velocity for different filters are reported with boxplot form in Fig.3-4, respectively. A filters box represents a point set of 100 MAEs of all Monte Carlo simulations, the characteristics of which can be reflected graphicly. The interquartile range (length of blue rectangle) and the distance between maximum and minimum value both indicate the dispersion degree of point set. Evidently, MAEs of CD-VCNF in 100 simulations is the most dense. Moreover, the medians (the red line in the blue rectangle) of CD-VCNF is obvious smaller than that of other filters. Hence, the estimation result of CD-VCNF is the most stable and accuracy one. In CD-VCNF, our proposed VSCM can can not only fit the state priori model, but also compensate the approximate error by VCPs. Through iteratively maximizing the ELBO by iteratively operating model identification and state correctness stages, VCPs can be identified so that the accuracy and adaptiveness of CD-VCNF can be improved gradually.

V. CONCLUSION

Based on VB framework, VCNF is proposed to solve filtering problem in continuous-discrete system. Through iteratively and alternatively operating model identification and

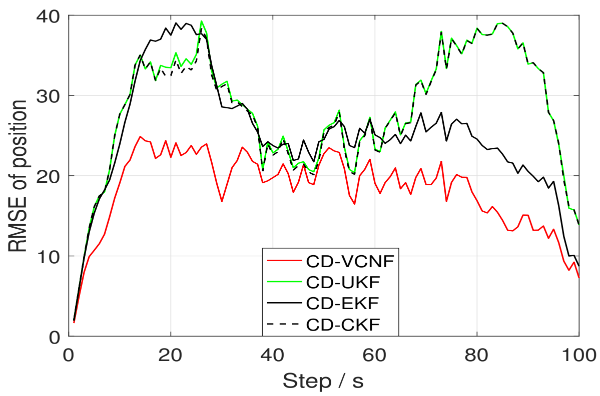

Fig. 1. RMSEs of position

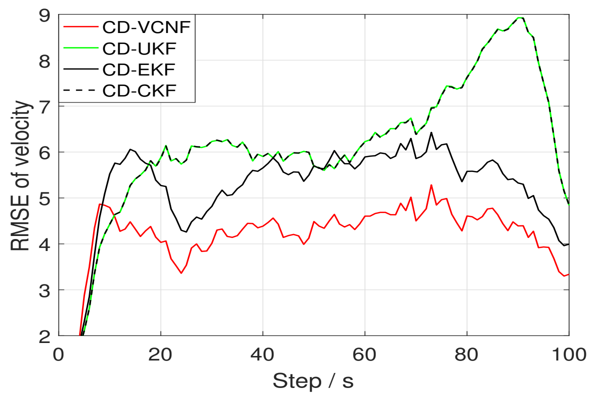

Fig. 2. RMSEs of velocity

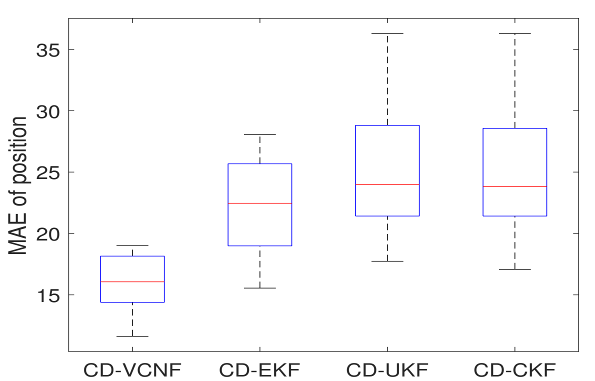

Fig. 3. MAEs of position

TABLE I
ARMSEs OF POSITION FOR DIFFERENT FILTERS

Filters	CD-VCNF	CD-EKF	CD-UKF	CD-CKF
ARMSEs(m)	18.8405	25.9768	29.1965	29.0043

TABLE II
ARMSEs OF VELOCITY FOR DIFFERENT FILTERS

Filters	CD-VCNF	CD-EKF	CD-UKF	CD-CKF
ARMSEs(m/s)	4.2569	5.3097	6.2960	6.2914

state correctness stages, ELBO can be maximized gradually, and at the same time, the identification of VCPs, the fitting of state priori model and the estimation of state can be achieved jointly. Through using VSCM to model uncertain state, performance of VCNF will not be influenced by uncertainty so that the adaptiveness and robustness to unpredictable error can be enhanced evidently. Moreover, because of a set of introduced VCPs in the mean and covariance of VSCM, the approximate and discretization error can be decreased and the estimation accuracy can be improved during the processing of the maximization of ELBO. Finally, we have shown that the performance of VCNF is superior in the simulation of target tracking.

ACKNOWLEDGMENT

This work was sponsored in part by the National Natural Science Foundation of China under Grants 61873208, 61203234, 61573287, 61135001, and 61374023, in part by the Shaanxi Natural Science Foundation of China under Grant 2017JM6006, in part by the Aviation Science Foundation of China under Grant 2016ZC3018, in part by the Fundamental Research Funds for Central Universities under Grant 3102017jghk02009, 3102018AX002 and in part by the Equipment Pre-research Foundation under Grant 2017-HT-XG.

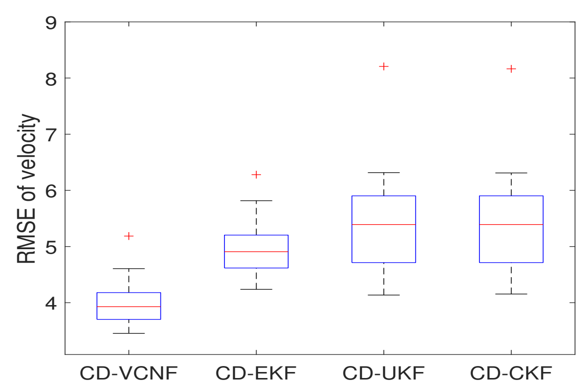

Fig. 4. MAEs of velocity

REFERENCES

[1] Y. Bar-Shalom, T. Kirubarajan, and X. R. Li, *Estimation with Applications to Tracking and Navigation*, 2001.

[2] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global Positioning Systems, Inertial Navigation, and Integration*, 2004.

[3] K. J. Astrom, *Introduction to stochastic control theory*, 1970.

[4] X. Liu and Z. Gao, "Robust finite-time fault estimation for stochastic nonlinear systems with brownian motions," *Journal of the Franklin Institute*, vol. 354, no. 6, pp. 2500–2523.

[5] WILKNISON and J. Darren, *Stochastic Modelling for Systems Biology*, 2006.

[6] A. H. Jazwinski, *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[7] P. Frogerais, J.-J. Bellanger, and L. Senhadji, "Various ways to compute the continuous-discrete extended kalman filter," *IEEE Transactions on Automatic Control*, vol. 57, no. 4, pp. 1000–1004, 2011.

[8] G. Y. Kulikov and M. V. Kulikova, "Estimating the state in stiff continuous-time stochastic systems within extended kalman filtering," *SIAM Journal on Scientific Computing*, vol. 38, no. 6, pp. A3565–A3588, 2016.

[9] M. V. Kulikova and G. Y. Kulikov, "Adaptive ode solvers in extended kalman filtering algorithms," *Journal of Computational and Applied Mathematics*, vol. 262, pp. 205–216, 2014.

[10] T. Knudsen and J. Leth, "A new continuous discrete unscented kalman filter," *IEEE Transactions on Automatic Control*, vol. 64, no. 5, pp. 2198–2205, 2018.

[11] I. Arasaratnam, S. Haykin, and T. R. Hurd, "Cubature kalman filtering for continuous-discrete systems: theory and simulations," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 4977–4993, 2010.

[12] S. Särkkä and A. Solin, "On continuous-discrete cubature kalman filtering," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 1221–1226, 2012.

[13] J. Wang, D. Zhang, and X. Shao, "New version of continuous–discrete cubature kalman filtering for nonlinear continuous–discrete systems," *ISA transactions*, 2019.

[14] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive kalman filtering by variational bayesian approximations," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, March 2009.

[15] G. Y. Kulikov and M. V. Kulikova, "Accurate numerical implementation of the continuous-discrete extended kalman filter," *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 273–279, 2013.

[16] ——, "The accurate continuous-discrete extended kalman filter for radar tracking," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 948–958, 2015.

[17] G. Y. Kulikov and M. V. Kulikova, "Accurate numerical implementation of the continuous-discrete extended kalman filter," *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 273–279, 2014.

[18] G. Y. Kulikov and M. V. Kulikova, "Accurate state estimation in continuous-discrete stochastic state-space systems with nonlinear or non-differentiable observations," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 4243–4250, 2017.

[19] M. Murata and K. Hiramatsu, "Non-gaussian filter for continuous-discrete models," *IEEE Transactions on Automatic Control*, 2019.

[20] Y. Xia, Z. Deng, L. Li, and X. Geng, "A new continuous-discrete particle filter for continuous-discrete nonlinear systems," *Information Sciences*, vol. 242, pp. 64–75, 2013.

[21] G. Y. Kulikov and M. V. Kulikova, "High-order accurate continuous-discrete extended kalman filter for chemical engineering," *European journal of Control*, vol. 21, pp. 14–26, 2015.

[22] O. Elerian, S. Chib, and N. Shephard, "Likelihood inference for discretely observed nonlinear diffusions," *Econometrica*, vol. 69, no. 4, pp. 959–993.

[23] C. M. Bishop, *Pattern recognition and machine learning*. Springer Science Business Media, 2006.