A minimum variance filter for continuous discrete systems with additive-multiplicative noise

Seham Allahyani
College of Engineering,
Design and Physical Sciences,
Brunel University, London, UK
Email:seham.allahyani@brunel.ac.uk

Paresh Date
College of Engineering,
Design and Physical Sciences,
Brunel University, London, UK
Email:paresh.date@brunel.ac.uk

***Abstract*—In this paper, we extend the minimum variance filter, which is proposed in the literature for discrete state space systems with multiplicative noise, to continuous-discrete systems with multiplicative noise. The differential equations that describe the process are discretised using the Euler scheme at a higher sampling frequency than the measurement frequency. We test the performance of our new filter *i.e.* continuous discrete filter (CDF) on simulated numerical examples and compare the results with discrete discrete filter (DDF) which ignores the state behaviour in-between the measurement samples. The results show that the CDF outperforms the DDF in all the cases examined.**

I. Introduction

The Bayesian filtering framework is a robust filtering framework where the state dynamics are typically modeled with stochastic differential or difference equations. In *continuous-discrete filtering*, the state dynamics is modelled as a continuous time system and the measurement equation is given in discrete time, *i.e.* the measurements (which are typically noisy) are available at discrete time instants. The measurement frequency may be limited by hardware or other physical considerations. The major difference between continuous discrete filter (henceforth abbreviated as CDF) and discrete time filter (or discrete discrete filter, abbreviated as DDF) is that, in the DDF approach, both the state dynamics and the noisy measurements are modeled as discrete-time processes. Filtering problems where a continuous-time signal is observed discretely in time have received a great deal of attention, since this kind of formulation often arise in numerous applications such as GPS and inertial navigation [1], stochastic control [2], target tracking [3] and finance [4]. The Bayesian optimal continuous discrete filter [5], [6] is the same as the discrete-time filter only when measurements are obtained at discrete time-instants, the posterior density is propagated from one sampling instant to the next by solving the associated Fokker-Planck equation.

In the literature, many conventional filtering algorithms are extended to deal with continuous-discrete systems. Example of such algorithms are Extended Kalman filter (EKF) [5] which approximates the exact solution by using a Taylor series expansion approximation to the nonlinear drift function and forms a Gaussian process approximation to the SDE, Particle filters (PF) [7] where a set of weighted particles is used for approximating the posterior probability measure and Unscented Kalman filter (UKF) [8] which relies on propagating a set of points representing a Gaussian density with the correct first two moments through the system equations. In [9], the extension of the cubature Kalman filter [10] to continuous-discrete filtering using Itô-Taylor expansion of continuous dynamics has been studied. The results in [11] which use the cubature integration method in continuous-discrete filtering were generalized in [12]. In [13], closed form solutions of continuous-discrete systems is derived. In [14], CDF algorithms using the extended Kalman filter (EKF), unscented Kalman filter (UKF), and particle filter (PF) with applications to the angle-only tracking in 3D are developed. Continuous-discrete filtering in a deterministic setting is discussed in [15].

Most of the results mentioned above are concerned with the additive noises only and multiplicative noise (or any other state-dependent noises) are not taken into account. A separate strand of research on the filtering problem for linear discrete time systems with multiplicative noise has also received a great deal of attention recently, since this kind of formulation has found many applications in sciences and engineering. Examples of such systems are encountered in signal processing systems, chemistry, economics, biological movement and ecology; see [16], [17], [18] and [19] and references therein. The second order statistics of the multiplicative noise, in contrast to the case of additive noise, is unknown. In an extended KF, multiplicative noise can act as a proxy for neglected higher order terms in Taylor series (since, unlike additive noise, it does depend on state). For systems subject to multiplicative noise, different kinds of algorithms have been introduced for the discrete time models. These algorithms are reported in [20], [21], [22], [23] and [16]. However, a similar problem was considered in [24] for continuous-discrete linear state space models where the solution is given in the form of solving a pair of coupled ODEs for the state estimate and its covariance matrix.

The motivation of this paper is to extend the result in [23] to deal with continuous discrete problems. Specifically, we consider a class of continuous discrete systems with additive as well as multiplicative noise, which includes square-root affine systems. In this paper, we use Euler scheme followed by conditional moment matching to transform stochastic differential equations (SDE) in the process equation into discrete model on a timescale which is finer than the measurement timescale. The problem addressed here is the design of a filter that minimizes the trace of estimation error covariance matrix at each measurement sampling instant. We demonstrate

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

through numerical experiments that our new filter performs better than the corresponding discrete discrete filter in [23], when information about continuous time dynamics is available.

The organization of the paper is as follows. In section 2, the aforementioned class of systems is described. We outline the problem in this section and derive its solution. In section 3, the proposed filtering algorithm is demonstrated by comprehensive numerical examples. Some concluding remarks given in section 4. Proofs of the main theorems in section 2 are provided in the Appendix.

II. Bayesian filtering algorithm under additive-multiplicative noise

A. State space model

Consider a system in which the process equation is described by a stochastic differential equation:

$$d\mathcal{X}(t) = (A\mathcal{X}(t) + B)dt + U_w d\mathcal{W}(t) + G\mathcal{X}^{\gamma}(t)d\mathcal{S}(t) \quad (1)$$

The behavior of the system is observed through noisy measurements $\mathcal{Y}(t_k)$ which are taken at discrete time instant $t_k = kT$, where T is the measurement sampling interval:

$$\mathcal{Y}(t_k) = C\mathcal{X}(t_k) + U_v\mathcal{V}(t_k). \quad (2)$$

Here, $\gamma \in \{0, 0.5, 1\}$, $\mathcal{X}(t)$ is an n-dimensional state of the system at any time t, $\mathcal{Y}(t_k) \in \mathbb{R}^r$ is the measurement at t_k^{th} time instant and A, B, G, C, D, U_w and U_v are given constant matrices of appropriate dimensions. $\mathcal{X}^{\gamma}(t)$ indicates a vector whose each element is the corresponding element of $\mathcal{X}(t)$ raised to the power γ. $\mathcal{W}(t) \in \mathbb{R}^n$ is a standard Wiener process with increment $d\mathcal{W}(t)$ and $\mathcal{V}(t_k), k = 1, 2, \cdots$ is a discrete time stochastic process which represents the measurement noise. The standard Wiener process $\mathcal{S}(t) \in \mathbb{R}^n$ represents the multiplicative noise. The initial state is a random vector with a known mean and covariance matrix, $\mathbb{E}[\mathcal{X}(0)] = \hat{\mathcal{X}}(0)$ and $\mathbb{E}[(\mathcal{X}(0) - \hat{\mathcal{X}}(0))(\mathcal{X}(0) - \hat{\mathcal{X}}(0))^{\top}] = P(0)$ respectively. $\mathcal{X}(0)$, $\mathcal{W}(t)$, $\mathcal{V}(t_k)$ and $\mathcal{S}(t)$ are mutually independent. This class of systems include systems with additive noise ($\gamma = 0$), multiplicative noise ($\gamma = 1$) and square root affine noise ($\gamma = 0.5$). The last case is especially relevant in financial mathematics; see, e.g. [25] and references therein.

The purpose of the optimal (Bayesian) continuous-discrete time filtering is to determine the evolution in time t of the conditional probability density function, also called the posterior density of the state defined for all $t \geq 0$:

$$p(\mathcal{X}(t)|\mathcal{Y}(t_1), ..., \mathcal{Y}(t_k)), \quad t \in [t_k, t_{k+1}), \quad k = 1, 2, ...$$

or at least the relevant moments of the distribution (e.g. the mean vector and the covariance matrix). The optimal continuous-discrete Bayesian filter is the same as the discrete filter performed in two steps.

1) Prediction step: In this step, the prior probability density function is evaluated by propagation of the previous posterior density between the measurement instants.

2) Update step: In this step, the posterior density is obtained by fusing with the predictive density using Bayesian rule. This step is the same as the discrete-time filter update step because the measurement-update relies only on the measurement equation, which is modeled in discrete time for a continuous-discrete state-space model case.

Solving the dynamic system (1)-(2) is very challenging since the SDEs appearing in the dynamic model or the corresponding Fokker-Planck-Kolmogorov partial differential equations cannot typically be solved analytically and approximation must be used. We outline our choice of discretisation scheme below. The discretisation is on a time-scale which is finer than that of measurement sampling time-scale.

B. Discretization of process model

Let $t_{k+1} - t_k = T$, $k > 0$ be the uniform time interval between consecutive measurement samples. Let $t_k^i \in [t_k, t_{k+1}], i = 1, 2, ..., m$ be uniformly spaced inter-sampling times, with $t_k^{i+1} - t_k^i =: \Delta = \frac{T}{m}$. Applying the Euler scheme to (1) over time interval $(t_k^i, t_k^i + \Delta)$ yields

$$\mathcal{X}(t_k^{i+1}) = \mathcal{X}(t_k^i) + (A\mathcal{X}(t_k^i) + B)\Delta + U_w\Delta W + \sum_{j=1}^{n}(G_j\,(\mathcal{X}_j^{\gamma}(t_k^i)))\Delta S_j.$$

where ΔW and ΔS are n-dimensional Gaussian random variables with zero mean and covariance matrices $\mathbb{E}[\Delta W \Delta W^{\top}] = \Delta$, $\mathbb{E}[\Delta S_j \Delta S_j^{\top}] = \Delta$ respectively. G_j represents j^{th} column of matrix G and ΔS_j is the j^{th} component of vector valued random variable ΔS.

In order to match exactly the first two moments, we use moment matching approach. So, the expression for the conditional mean of $\mathcal{X}(t_k^{i+1})$ given $\mathcal{X}(t_k^i)$ can be easily shown to be

$$\mathbb{E}[\mathcal{X}(t_k^{i+1})|\mathcal{X}(t_k^i)] = \mathcal{X}(t_k^i) + (A\mathcal{X}(t_k^i) + B))\Delta,$$

where $i = 1, 2.., m-1$ and the associated conditional covariance matrix is

$$var[\mathcal{X}(t_k^{i+1})|\mathcal{X}(t_k^i)] = U_w U_w^{\top}\Delta + \sum_{j=1}^{n}(G_j G_j^{\top}\,(\mathcal{X}_j^{2\gamma}(t_k^i)))\Delta$$

Then

$$\mathcal{X}(t_k^{i+1}) = \tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m\tilde{\mathcal{W}}(t_k^{i+1}) + \tilde{G}\,\mathrm{diag}(\mathcal{X}_j^{\gamma}(t_k^i))\tilde{\mathcal{S}}(t_k^{i+1}) \quad (3)$$

where

$$\tilde{A} = I + A\Delta,\ \tilde{B} = B\Delta,$$
$$\tilde{U}_m = U_w\sqrt{\Delta},\ \tilde{G} = G\sqrt{\Delta},$$

and $\tilde{\mathcal{W}}$, $\tilde{\mathcal{S}}$ are uncorrelated, zero mean random processes with identity covariance matrices. This puts the system in a discrete state space framework for which the standard discrete time filtering tools can be applied.

C. Minimum variance filter for the continuous-discrete system

To derive the recursive filtering equations, it is assumed that the observations are given up to time t_k and that the approximate conditional mean of $\mathcal{X}(t_k^i)$ given $\mathcal{Y}(t_k)$, $\hat{\mathcal{X}}(t_k^i|t_k)$, is available. From this value, the approximated conditional mean of $\mathcal{X}(t_k^{i+1})$, which provides the predictor, $\hat{\mathcal{X}}(t_k^{i+1}|t_k)$, is derived using (3):

$$\hat{\mathcal{X}}(t_k^{i+1}|t_k) = \tilde{A}\hat{\mathcal{X}}(t_k^i|t_k) + \tilde{B}. \tag{4}$$

The predicted estimate $\hat{\mathcal{X}}(t_k^{i+1}|t_k)$ needs to be updated with the information provided by $\mathcal{Y}(t_{k+1})$, to obtain the filtered estimate. So, the update equation for a linear filter is

$$\hat{\mathcal{X}}(t_k^{i+1}|t_{k+1}) = \hat{\mathcal{X}}(t_k^{i+1}|t_k) + \bar{K}(k+1)(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k)), \tag{5}$$

where $\hat{\mathcal{X}}(t_k^{i+1}|t_{k+1})$ indicates updated estimate of $\hat{\mathcal{X}}(t_k^{i+1})$ after $\mathcal{Y}(t_{k+1})$ becomes available, and the covariance matrix is given by

$$\begin{aligned}\bar{P}(t_k^{i+1}|t_k) &= \mathbb{E}[(\mathcal{X}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^{i+1}|t_k))\\ &((\mathcal{X}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^{i+1}|t_k))^\top].\end{aligned} \tag{6}$$

As in [23], our objective is to find a filter gain $\bar{K}(k+1)$ that would minimize the trace of the covariance $\bar{P}(t_k^{i+1}|t_k)$ of the state estimate $\hat{\mathcal{X}}(t_k^{i+1}|t_{k+1})$ and obtain an expression for the optimum filter. Our main result in this section, which is an extension of the corresponding result from [23], is given in the next theorem.

Theorem 1: For equation (2) and (3), the filter gain $\bar{K}(k+1)$ that minimizes the trace of the covariance $\bar{P}(t_k^{i+1}|t_k)$ is given by

$$\bar{K}(k+1) = \bar{P}(t_{k+1}^i|t_k^i)C^\top[C\bar{P}(t_{k+1}^i|t_k^i)C^\top + U_vU_v^\top]^{-1}, \tag{7}$$

where

$$\begin{aligned}\bar{P}(t_{k+1}^i|t_k^i) &= \tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top + \tilde{U}_m\tilde{U}_m^\top +\\ &\tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top\end{aligned} \tag{8}$$

and

$$\begin{aligned}&\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\\ &= \operatorname{diag}(\bar{P}_{jj}(t_k^i|t_k^i) + (\hat{\mathcal{X}}_j(t_k^i|t_k^i))^2)) \quad \text{if} \gamma = 1,\\ &= \operatorname{diag}(\hat{\mathcal{X}}_j(t_k^i|t_k^i)) \quad \text{if} \quad \gamma = 0.5,\\ &= \operatorname{diag}(I_n) \quad \text{if} \quad \gamma = 0.\end{aligned} \tag{9}$$

Proof: See Appendix.
A few of remarks on this result are in order.

- If $G = 0$ i.e. if there is no multiplicative noise, our filter reduces to the Kalman Bucy filter for continuous-discrete models with additive noise [26].
- The major difference between our work and the work represented in [23] is that, in this paper the dynamics are modeled as continuous-time process and the measurements are modeled as discrete-time process while in [23] both the dynamics and measurements are modeled as discrete-time processes. This requires updating values of $\mathcal{X}(t)$ at $t \in (t_k, t_{k+1})$. As will be seen in the numerical examples, this improves the prediction quality. If $m = 1$, we recover the results from [23].
- As mentioned earlier, the three specific values of γ viz. $0, 0.5$ and 1 encompass the cases of additive, square root affine and multiplicative noise respectively. As will be seen later in the proof (see equation (9), in particular), these choices of γ still allow us to get a closed-form recursive expression for the covariance matrix. Further, note that $\gamma \leq 1$ in our set up is sufficient for the Euler scheme to converge; see ([17], chapter 10), for example.

III. Numerical example

A. Example 1

For numerically evaluation, we used the same model parameters was described in [23] for the system (1)- (2) which are:
$A = -0.5$, $B = 0$, $C = 10$, $U_w = 6$, $U_v = 1$, and $G = 0.1$. $\mathcal{W}(t)$ and $\mathcal{S}(t)$ are standard Wiener process, $\mathcal{V}(t_k)$ is random variable with with $\mathbb{E}(\mathcal{V}(t_k)) = 0$ and $\mathbb{E}(\mathcal{V}(t_k))^2 = 1$ and uncorrelated with $\mathcal{W}(t)$ and $\mathcal{S}(t)$. Initial conditions are $\mathcal{X}(0) = 1$, $\hat{\mathcal{X}}(0) = 0$ and $P(0) = 1$.
The measurement sampling period $T = 1$. We consider a sequence of two different time steps between $t_k = kT$ and $t_{k+1} = (k+1)T$ $m = 5$ and $m = 10$, so $\Delta = 1/5$ and $\Delta = 1/10$. Then we will use these parameters to derive the discretization parameters presented in the Section 2.2. This gives

$$\begin{aligned}&\tilde{A} = 0.9, \quad \tilde{B} = 0, \quad \tilde{U}_w = 2.6833, \quad \tilde{G} = 0.0447,\\ &\tilde{A} = 0.95, \quad \tilde{B} = 0, \quad \tilde{U}_w = 1.8974, \quad \tilde{G} = 0.0316.\end{aligned}$$

for $\Delta = 1/5$ and $\Delta = 1/10$, respectively. In order to compare the performance of the estimators, we use the root mean square error (RMSE) and the mean relative absolute error (MRAE) criteria. Consider 100 independent simulations, each with 200 data points. Denoting $\mathcal{Y}^{(s)}(t_k)$, $k = 1, ..., 200$ as the s^{th} set of true values of the measurement, and $\hat{\mathcal{Y}}^{(s)}(t_k|t_k)$ as the filtered measurement estimate at time t_k for the s^{th} simulation run, the RMSE and MRAE of the filter for each of the algorithms is calculated by

$$RMSE(s) = \sqrt{\frac{1}{200}\sum_{t_k=1}^{200}(\mathcal{Y}^{(s)}(t_k) - \hat{\mathcal{Y}}^{(s)}(t_k|t_k))^2}, \quad s = 1, ..., 100,$$

$$MRAE(s) = \frac{1}{200}\sum_{t_k=1}^{200}\frac{|\mathcal{Y}^{(s)}(t_k) - \hat{\mathcal{Y}}^{(s)}(t_k|t_k)|}{|\mathcal{Y}^{(s)}(t_k)|}, \quad s = 1, ..., 100.$$

Then the average of RMSE and MRAE for each of the states over 100 simulations is given by

$$AvRMSE = \frac{1}{100}\sum_{s=1}^{100} RMSE(s).$$

$$AvMRAE = \frac{1}{100}\sum_{s=1}^{100} RMSE(s).$$

Using our estimator, we will compare the performance of two filters: continuous-discrete filter and discrete discrete filter *i.e.* filter presented in [23] with different values of γ . The results of continuous-discrete filter and discrete discrete filter will be represented by CDF and DDF, respectively. The results are summarized in Tables I and II. We can see from these tables that, in all the cases, the estimators with CDF perform better than the estimator with DDF.

TABLE I. COMPARISON OF $AvRMSE$ AND $AvMRAE$ FOR CDF AND DDF FOR DIFFERENT VALUES OF γ AND WITH $\Delta = 1/5$

		$\gamma = 0$	$\gamma = 0.5$	$\gamma = 1$
$AvRMSE$	CDF	0.9850	0.9864	0.9044
	DDF	1.4572	1.4794	1.4664
$AvMRAE$	CDF	0.9800	0.9691	0.9700
	DDF	1.6485	1.4573	1.4548

TABLE II. COMPARISON OF $AvRMSE$ AND $AvMRAE$ FOR CDF AND DDF FOR DIFFERENT VALUES OF γ AND WITH $\Delta = 1/10$

		$\gamma = 0$	$\gamma = 0.5$	$\gamma = 1$
$AvRMSE$	CDF	0.9758	0.9865	0.9801
	DDF	1.4635	1.4796	1.4699
$AvMRAE$	CDF	0.9625	0.9801	0.9700
	DDF	1.5415	1.4782	1.4548

B. Example 2

As another example, we consider the same model parameters was described in [23] for the system (1)- (2) which are:

$$A = \begin{bmatrix} -1 & 0.5 \\ 1 & -1 \end{bmatrix}, \quad B = 0 \quad U_w = \begin{bmatrix} -6 \\ 1 \end{bmatrix},$$
$$C = \begin{bmatrix} -100 \\ 10 \end{bmatrix}, \quad G = \begin{bmatrix} 0.12 & 0.02 \\ 0.15 & 0.1 \end{bmatrix},$$

and $U_v = 1$. As before, $\mathcal{W}(t)$ and $\mathcal{S}(t)$ are standard Wiener process, $\mathcal{V}(t_k)$ is random variable with zero mean and identity covariance matrix $\mathcal{I}$ and uncorrelated with $\mathcal{W}(t)$ and $\mathcal{S}(t)$. The initial conditions are:
$\mathcal{X}(0) = [\,0 \quad 1\,]^\top$, $\hat{\mathcal{X}}(0) = [\,0 \quad 0\,]^\top$ and $P(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
The measurement sampling period $T = 1$. We consider a sequence of 10 time steps between $t_k = kT$ and $t_{k+1} = (k+1)T$, so $\Delta = 1/10$. A difference in our simulation was that instead of using the parameters described in [23] themselves we used them to derive the discretization parameters presented in the Section 2.2. That is,

$$\tilde{A} = \begin{bmatrix} 0.9 & -0.05 \\ 0.1 & 0.9 \end{bmatrix}, \quad \tilde{B} = 0 \quad \tilde{U}_w = \begin{bmatrix} -1.8974 \\ 0.3162 \end{bmatrix},$$
$$\tilde{G} = \begin{bmatrix} 0.0379 & 0.0063 \\ 0.0474 & 0.0316 \end{bmatrix},$$

We will compare the performance of two filters for this data generating system: continuous-discrete filter and discrete discrete filter *i.e.* filter presented in [23] with different values of γ . In keeping with the notation in the previous subsection, the results of the continuous-discrete filter and discrete discrete filter will be represented by CDF and DDF, respectively. The AvRMSE and AvMRAE are calculated for these filters with different values of γ as the previous subsection.

Table III summarizes the results of this experiment. As can be seen, the CDF provides better accuracy when compared to DDF in all the cases.

TABLE III. COMPARISON OF $AvRMSE$ AND $AvMRAE$ FOR CDF AND DDF FOR DIFFERENT VALUES OF γ

		$\gamma = 0$	$\gamma = 0.5$	$\gamma = 1$
$AvRMSE$	CDF	0.9622	0.9798	0.9634
	DDF	2.5929	2.6408	2.5947
$AvMRAE$	CDF	0.9600	0.9402	0.9400
	DDF	2.5315	2.5314	2.5231

IV. CONCLUSION

In this paper, the optimal linear minimum variance filter is derived for a class of continuous-discrete time systems with additive as well as multiplicative noise. The closed form solution generalizes the results for minimum variance filtering for additive-multiplicative noise case in [23]. Discretization of the continuous-time dynamic model using the Euler scheme is described. The results of this paper were applied to simulated linear system with additive-multiplicative noises. Our numerical experiments indicate that the continuous-discrete filter outperforms discrete-discrete filter.

APPENDIX

The filtering estimates of the state covariance is obtained by combining the equations (2)-(5) as follows. For brevity of notation, an expression $LL^\top$ will sometimes be denoted as $(L)(\star)^\top$, where L is a matrix-valued expression and where there is no risk of confusion. The state covariance matrix at time $k+1$ can be written as

$$\begin{aligned}
&\bar{P}(t_k^{i+1}|t_k) = \mathbb{E}[(\mathcal{X}(t_k^{i+1}) - \mathcal{X}(t_k^{i+1}|t_k))(\star)^\top] \\
&= \mathbb{E}[\tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m\mathcal{W}(t_k^{i+1}) + \tilde{G}\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^{i+1}) - \\
&(\tilde{A}\hat{\mathcal{X}}(t_k^i|t_k) + \tilde{B} + \bar{K}(k+1)(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k)))(\star)^\top] \\
&= \mathbb{E}[\tilde{A}(\mathcal{X}(t_k^i) - \hat{\mathcal{X}}(t_k^i|t_k))(\star)^\top] + \tilde{U}_m\tilde{U}_m^\top \\
&+ \tilde{G}\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top + \\
&\bar{K}(k+1)(\mathbb{E}(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k))(\star)^\top)\bar{K}(k+1)^\top - \\
&(\tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m\mathcal{W}(t_k^{i+1}) + \tilde{G}\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^{i+1}) \\
&- \hat{\mathcal{X}}(t_k^i|t_k))(\bar{K}(k+1)(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k)))^\top - \\
&(\bar{K}(k+1)(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k)))(\tilde{A}\mathcal{X}(t_k^i) + \tilde{B} + \tilde{U}_m\mathcal{W}(t_k^{i+1}) \\
&+ \tilde{G}\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^{i+1}) - \hat{\mathcal{X}}(t_k^i|t_k))^\top \qquad (10)
\end{aligned}$$

Next, we need the following covariance term in evaluating $\bar{P}(t_k^{i+1}|t_k)$:

$$\begin{aligned}
&\mathbb{E}[(\mathcal{Y}(t_{k+1}) - \hat{\mathcal{Y}}(t_{k+1}|t_k))(\star)^\top)] = \\
&C\tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top C^\top + C\tilde{U}_m\tilde{U}_m^\top C^\top + \\
&C\tilde{G}(\mathbb{E}[\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))\operatorname{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top])\tilde{G}^\top C^\top + U_vU_v^\top \qquad (11)
\end{aligned}$$

We also need to evaluate some cross covariance terms, whose expressions are derived next:

$$\begin{aligned}&\mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i)-\hat{\mathcal{X}}(t_k^i|t_k))+\tilde{U}_m\mathcal{W}(t_k^{i+1})+\\&\tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^{i+1}))(\mathcal{Y}(t_{k+1})-\hat{\mathcal{Y}}(t_{k+1}|t_k))^\top]\\&=\mathbb{E}[(\tilde{A}(\mathcal{X}(t_k^i)-\hat{\mathcal{X}}(t_k^i|t_k))+\tilde{U}_m\mathcal{W}(t_k^{i+1})+\\&\tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^{i+1}))(C(\tilde{A}(\mathcal{X}(t_k^i)-\hat{\mathcal{X}}(t_k^i|t_k))+\\&\tilde{U}_m\mathcal{W}(t_k^{i+1})+\tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^{i+1}))+U_v\mathcal{V}(t_{k+1})\\&=\tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top C^\top+\tilde{U}_m\tilde{U}_m^\top C^\top+\\&\tilde{G}\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top C^\top\end{aligned}\tag{12}$$

Further,

$$\begin{aligned}&\mathbb{E}[(\mathcal{Y}(t_{k+1})-\hat{\mathcal{Y}}(t_{k+1}|t_k))(\tilde{A}(\mathcal{X}(t_k^i)-\hat{\mathcal{X}}(t_k^i|t_k))+\tilde{U}_m\mathcal{W}(t_k^{i+1})\\&+\tilde{G}\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\mathcal{S}(t_k^{i+1}))^\top]\\&=C\tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top+C\tilde{U}_m\tilde{U}_m^\top+C\\&\tilde{G}\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top\end{aligned}\tag{13}$$

Substituting (11), (12) and (13) in (10), we have

$$\begin{aligned}&\bar{P}(t_k^{i+1}|t_k)=\tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top+\tilde{U}_m\tilde{U}_m^\top+\\&\tilde{G}\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top\\&+\bar{K}(k+1)(C\tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top C^\top+C\tilde{U}_m\tilde{U}_m^\top C^\top+C\tilde{G}\\&(\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top])\tilde{G}^\top C^\top+U_vU_v^\top)\bar{K}(k+1)^\top\\&-2\bar{K}(k+1)(\tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top C^\top+\tilde{U}_m\tilde{U}_m^\top C^\top+\\&\tilde{G}\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top C^\top)\\&=\bar{P}(t_{k+1}^i|t_k^i)+\bar{K}(k+1)(C\bar{P}(t_{k+1}^i|t_k^i)C^\top+U_vU_v^\top)\bar{K}(k+1)^\top\\&-2\bar{K}(k+1)\bar{P}(t_{k+1}^i|t_k^i)C^\top,\end{aligned}$$

where

$$\begin{aligned}&\bar{P}(t_{k+1}^i|t_k^i)=\tilde{A}\bar{P}(t_k^i|t_k)\tilde{A}^\top+\tilde{U}_m\tilde{U}_m^\top+\\&\tilde{G}\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]\tilde{G}^\top,\end{aligned}\tag{14}$$

and $\mathbb{E}[\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))\,\mathrm{diag}(\mathcal{X}_j^\gamma(t_k^i))^\top]$ is as in equation (9). To find the value of $\bar{K}(k+1)$ that minimizes the trace of the covariance $\bar{P}(t_k^{i+1}|t_k)$ we differentiate the trace of the above expression with respect to the filter gain matrix $\bar{K}(k+1)$ and set the derivative to zero.

$$\begin{aligned}&\frac{\partial tr\,\bar{P}(t_k^{i+1}|t_k)}{\partial\bar{K}(k+1)}=-2\bar{P}(t_{k+1}^i|t_k^i)C^\top+2\bar{K}(k+1)\\&[C\bar{P}(t_{k+1}^i|t_k^i)C^\top+U_vU_v^\top].\end{aligned}\tag{15}$$

Setting this partial derivative to zero leads the following expression for $\bar{K}(k+1)$:

$$\bar{K}(k+1)=\bar{P}(t_{k+1}^i|t_k^i)C^\top[C\bar{P}(t_{k+1}^i|t_k^i)C^\top+U_vU_v^\top]^{-1}.\tag{16}$$

which is the required result. ∎

References

[1] Y. Bar-Shalom, X. R. Li, T. Kirubarajan, Estimation with applications to tracking and navigation: theory algorithms and software, John Wiley & Sons, 2004.

[2] K. J. Åström, Introduction to stochastic control theory, Courier Corporation, 2012.

[3] S. Blackrnan, A. House, Design and analysis of modern tracking systems, Boston, MA: Artech House.

[4] B. Oksendal, Stochastic differential equations: an introduction with applications, Springer Science & Business Media, 2013.

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

[6] P. S. Maybeck, Stochastic models, estimation, and control, Vol. 2, Academic press, 1982.

[7] A. Doucet, N. de Freitas, N. Gordon, Sequential monte carlo methods in practice, Statistics.

[8] S. Särkkä, On unscented Kalman filtering for state estimation of continuous-time nonlinear systems, Automatic Control, IEEE Transactions on 52 (9) (2007) 1631–1641.

[9] I. Arasaratnam, S. Haykin, T. R. Hurd, Cubature Kalman filtering for continuous-discrete systems: theory and simulations, Signal Processing, IEEE Transactions on 58 (10) (2010) 4977–4993.

[10] I. Arasaratnam, S. Haykin, Cubature Kalman filters, Automatic Control, IEEE Transactions on 54 (6) (2009) 1254–1269.

[11] S. Särkkä, A. Solin, On continuous-discrete cubature Kalman filtering, in: Proceedings of SYSID, Vol. 2012, 2012.

[12] S. Särkkä, J. Sarmavuori, Gaussian filtering and smoothing for continuous-discrete dynamic systems, Signal Processing 93 (2) (2013) 500–510.

[13] F. E. Daum, Exact finite dimensional nonlinear filters for continuous time processes with discrete time measurements, in: Decision and Control, 1984. The 23rd IEEE Conference on, IEEE, 1984, pp. 16–22.

[14] M. Mallick, M. Morelande, L. Mihaylova, Continuous-discrete filtering using EKF, UKF, and PF, in: Information Fusion (FUSION), 2012 15th International Conference on, IEEE, 2012, pp. 1087–1094.

[15] Y. Yamamoto, M. Nagahara, P. P. Khargonekar, Signal reconstruction via sampled-data control theory: Beyond the Shannon paradigm, Signal Processing, IEEE Transactions on 60 (2) (2012) 613–625.

[16] E. Gershon, U. Shaked, I. Yaesh, H_∞-control and filtering of discrete-time stochastic systems with multiplicative noise, Automatica 37 (3) (2001) 409–417.

[17] P. E. Kloeden, E. Platen, Numerical solution of stochastic differential equations, Vol. 23, Springer, 1992.

[18] K. Jordan, A minimal intervention principle for coordinated movement, in: Proceedings of the 2002 Conference on Advances in Neural Information Processing Systems, Vol. 15, The MIT Press Cambridge, MA, 2003, p. 27.

[19] R. R. Mohler, Bilinear control processes: with applications to engineering, ecology and medicine, Academic Press, Inc., 1973.

[20] W. Li, E. Todorov, R. E. Skelton, Estimation and control of systems with multiplicative noise via linear matrix inequalities, in: the American Control Conference, Proceedings of, IEEE, 2005, pp. 1811–1816.

[21] F. Carravetta, A. Germani, N. Raimondi, Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise, Automatic Control, IEEE Transactions on 42 (8) (1997) 1106–1126.

[22] O. L. Costa, G. R. Benites, Linear minimum mean square filter for discrete-time linear systems with markov jumps and multiplicative noises, Automatica 47 (3) (2011) 466–476.

[23] K. Ponomareva, P. Date, An exact minimum variance filter for a class of discrete time systems with random parameter perturbations, Applied Mathematical Modelling 38 (9) (2014) 2422–2434.

[24] J. Jimenez, T. Ozaki, Linear estimation of continuous-discrete linear state space models with multiplicative noise, Systems & control letters 47 (2) (2002) 91–101.

[25] J. C. Cox, J. E. Ingersoll, S. A. Ross, A theory of the term structure of interest rates, Econometrica 53 (2) (1985) 385–407.

[26] R. E. Kalman, A new approach to linear filtering and prediction problems, Transactions of AMSE (Journal of Basic Engineering) 82 D (1) (1960) 35–45.