

A minimum variance filter for discrete time linear systems with parametric uncertainty

Seham Allahyani and Paresh Date

Abstract—A minimum variance filter for a class of discrete time systems with additive as well as multiplicative noise is investigated in this paper. We extend the results from recent work by Ponomareva and Date to account for multiplicative noise in the measurement equation. More importantly, we provide an interpretation of the multiplicative noise in both transition and measurement equations in terms of parameter perturbations in a linear additive model. The utility of the proposed filtering algorithm is demonstrated through numerical simulation experiments using models from academic literature where the parameters are estimated from real data.

I. INTRODUCTION

Estimation and filtering in noise is an important problem in many practical areas, including economics, tracking, weather forecasting, navigation systems, control and signal processing. The additive noise state space model has received much attention in the filtering literature (see, for instance, [1], [2], [3], [4]). On the other hand, in many practical systems we need to consider the noise component to be both multiplicative and additive to the signal component. Compared to the additive noise case, the corresponding filtering problem for systems with multiplicative noise has gained somewhat less attention. Multiplicative noise has been observed in many applications in sciences and engineering such as signal processing systems, chemistry, economics, biological movement and ecology; see [5], [6], [7], [8] and references therein. The second order statistics of the multiplicative noise, in contrast to the case of the additive noise, is unknown *a priori*. However, multiplicative noise model can be used to model the stochastic uncertainty in the system parameters which are estimated from data, and a demonstration of this fact is one of the contributions in the paper. In an extended KF, multiplicative noise can act as a proxy for neglected higher order terms in Taylor series (since, unlike the additive noise, it does depend on the state). So far, discrete time filtering problems subject to multiplicative noise has been studied recently in several papers. This includes a filtering algorithm which uses linear matrix inequalities to guarantee that the covariance error is bounded from above by a specified positive definite matrix [9], an optimal filter within a class of polynomial transformations of a fixed degree [10], a linear minimum mean square estimator (LMMSE) subject to state and measurement multiplicative noises and Markov jumps in the parameter vector [11], a robust finite-horizon Kalman filter [12] and a minimum variance linear filter for a class of

systems which includes multiplicative noise [13]. In [14], a new structure of a linear recursive estimator minimizing the mean square error is derived for a system with multiplicative noise in the measurement model. The results in [14] were generalized in [15] to develop a different structure of linear recursive estimator for a nonzero mean signal corrupted by multiplicative noise. In [5] the filtering and control problem under the H_∞ criterion has been studied. A similar problem was considered in [16] for continuous-discrete linear state space models.

In this paper we consider an extension of Ponomareva and Date's work in [13] on filtering in systems with multiplicative noise. Specifically, we extend the results from [13] to propose a complete, closed-form solution to the minimum variance filtering problem for linear systems with multiplicative noise in both transition and measurement equations and demonstrate its performance through numerical simulation experiments. The main contribution of the paper is a demonstration of how filtering under multiplicative noise can act as a proxy for filtering under parameter uncertainty, which is characterized as random perturbations of the state space matrices. A limited amount of work has been done on minimum variance filtering under parametric uncertainty, although a related robust estimation problem under parametric uncertainty has been recently addressed in [17].

The remaining paper is organized as follows. In section II, the aforementioned class of systems is described. We outline the problem and derive its solution in this section. In section III, the proposed filtering algorithm is demonstrated by comprehensive numerical examples. Some concluding remarks given in section IV. Proof of the main theorem in section II is provided in the Appendix.

II. SYSTEM MODEL AND PROBLEM FORMULATION

The system dynamics under consideration can be described by the following discrete time equation:

$$\mathcal{X}(k+1) = A\mathcal{X}(k) + B + U_w\mathcal{W}(k) + G_1 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_1(k), \quad (1)$$

while the measurement model is

$$\mathcal{Y}(k) = C\mathcal{X}(k) + D + U_v\mathcal{V}(k) + G_2 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_2(k). \quad (2)$$

Here $\mathcal{X}(k) \in \mathbb{R}^n$ is the state vector at time k , $\mathcal{Y}(k) \in \mathbb{R}^r$ is the measurement vector at time k and $A, B, G_1, C, D, G_2, U_w$ and U_v are given deterministic matrices. For a vector $\mathcal{Z}, M = \text{diag}(\mathcal{Z}_j)$ represents a diagonal matrix with $M_{jj} =$

\mathcal{Z}_j . $\mathcal{W}(k) \in \mathbb{R}^n$ and $\mathcal{V}(k) \in \mathbb{R}^r$ are the process noise and the measurement noise, respectively. The random variables $\mathcal{S}_1(k) \in \mathbb{R}^n$ and $\mathcal{S}_2(k) \in \mathbb{R}^r$ represent the multiplicative noise sources. Note that (1) and (2) contain both additive as well as multiplicative noise. We make two assumptions:

- 1) The noise signals $\mathcal{W}(k)$, $\mathcal{V}(k)$, $\mathcal{S}_1(k)$ and $\mathcal{S}_2(k)$ are zero mean, i.i.d. random vectors with identity covariance matrix \mathcal{I} .
- 2) 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. $\mathcal{X}(0)$, $\mathcal{W}(k)$, $\mathcal{V}(k)$, $\mathcal{S}_1(k)$ and $\mathcal{S}_2(k)$ are mutually independent.

Under the above assumptions, note that one may also treat $\mathcal{S}_1(k)$ and $\mathcal{S}_2(k)$ as zero mean random perturbations in the system matrices A and C which have covariance matrices $G_1 G_1^\top$ and $G_2 G_2^\top$ respectively. This can be seen by rearranging (1) and (2) as

$$\mathcal{X}(k+1) = (A + G_1 \text{diag}(\mathcal{S}_{1j}(k))) \mathcal{X}(k) + B + U_w \mathcal{W}(k), \quad (3)$$

$$\mathcal{Y}(k) = (C + G_2 \text{diag}(\mathcal{S}_{2j}(k))) \mathcal{X}(k) + D + U_v \mathcal{V}(k). \quad (4)$$

We will later use this interpretation in the numerical experiments in section III. Assume that the observations up to time k are given and that approximations of the conditional mean of $\mathcal{X}(k)$ given $\mathcal{Y}(k)$, $\hat{\mathcal{X}}(k|k)$, is available. From this value, the approximated conditional mean of $\mathcal{X}(k+1)$, which provides the predictor, $\hat{\mathcal{X}}(k+1|k)$, is derived using (1):

$$\hat{\mathcal{X}}(k+1|k) = A \hat{\mathcal{X}}(k|k) + B. \quad (5)$$

The predictor $\hat{\mathcal{X}}(k+1|k)$ must now be updated with the information provided by the new measurement, $\mathcal{Y}(k)$, to obtain the filter. The update equation for a linear filter is

$$\begin{aligned} \hat{\mathcal{X}}(k+1|k+1) = \\ \hat{\mathcal{X}}(k+1|k) + \bar{K}(k+1)(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k)), \end{aligned} \quad (6)$$

and the covariance matrix is given by

$$\bar{P}(k+1|k+1) = \mathbb{E}[\Phi(k+1)\Phi(k+1)^\top], \quad (7)$$

where $\Phi(k) := \mathcal{X}(k+1) - \hat{\mathcal{X}}(k+1|k+1)$. Our objective is to find a filter gain $\bar{K}(k+1)$ that would minimize the trace of the covariance $\bar{P}(k+1|k+1)$ of the state estimate $\hat{\mathcal{X}}(k+1|k+1)$ and obtain an expression for the optimum filter. Our main result in this section, which is an extension of the corresponding result from [13], is given in the next theorem.

Theorem 1: For system (1)-(2) with assumptions 1-2, the filter gain $\bar{K}(k+1)$ that minimizes the trace of the covariance $\bar{P}(k+1|k+1)$ is given by

$$\begin{aligned} \bar{K}(k+1) = \bar{P}(k+1|k)C^\top [C\bar{P}(k+1|k)C^\top + U_v U_v^\top + \\ G_2(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_2^\top]^{-1} \end{aligned} \quad (8)$$

where

$$\begin{aligned} \bar{P}(k+1|k) = A\bar{P}(k|k)A^\top + U_w U_w^\top + \\ G_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top. \end{aligned} \quad (9)$$

Proof: See Appendix.

Remark 1: If $G_2 = 0$, our filter reduces to a special case of the filter with state multiplicative noise only, which was discussed earlier in [13], with $\gamma = 1$ in the authors' notation in that paper. If, in addition, $G_1 = 0$ i.e. if there is no multiplicative noise either in the transition equation or in the measurement equation, our filter reduces to the Kalman filter for the linear additive noise case.

III. NUMERICAL EXAMPLES

We consider two numerical examples in this section. In both the cases, the parameters of the state space model under consideration are estimated from real data in the literature. The purpose of these examples is to demonstrate how one can use the information about randomness in the parameters estimated from data (e.g., in terms of the asymptotic variance of the parameter estimates) to design a filter which minimizes variance relative to the additive noise as well as the *noise* introduced by parameter uncertainty.

A. Example 1

In this example, we consider a two factor extension of the Vasicek interest rate model [18]. The treatment below and the estimated parameter values are from [19], although the model has been treated quite extensively in the econometric and financial literature; see, e.g. [20] and [21] for empirical studies, among others. The transition equation in the state space formulation of this model is expressed as the following:

$$\begin{aligned} \begin{bmatrix} \mathcal{X}_1(k+1) \\ \mathcal{X}_2(k+1) \end{bmatrix} = \underbrace{\begin{bmatrix} e^{-\kappa_1 \Delta t} & 0 \\ 0 & e^{-\kappa_2 \Delta t} \end{bmatrix}}_A \begin{bmatrix} \mathcal{X}_1(k) \\ \mathcal{X}_2(k) \end{bmatrix} + \\ \begin{bmatrix} \theta_1(1 - e^{-\kappa_1 \Delta t}) \\ \theta_2(1 - e^{-\kappa_2 \Delta t}) \end{bmatrix} + \begin{bmatrix} \mathcal{W}_1(k+1) \\ \mathcal{W}_2(k+1) \end{bmatrix}, \end{aligned} \quad (10)$$

where $\mathcal{W}(k+1) = [\mathcal{W}_1(k+1) \quad \mathcal{W}_2(k+1)]^\top \sim N(0, W)$, with

$$W = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1}(1 - e^{-2\kappa_1 \Delta t}) & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2}(1 - e^{-2\kappa_2 \Delta t}) \end{bmatrix},$$

and the measurement equation is:

$$\begin{bmatrix} \mathcal{Y}(t_k, T_1) \\ \mathcal{Y}(t_k, T_2) \\ \vdots \\ \mathcal{Y}_1(t_k, T_m) \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{F_1(t_k, T_1)}{T_1 - t_k} & \frac{F_2(t_k, T_1)}{T_1 - t_k} \\ \frac{F_1(t_k, T_2)}{T_2 - t_k} & \frac{F_2(t_k, T_2)}{T_2 - t_k} \\ \vdots & \vdots \\ \frac{F_1(t_k, T_m)}{T_m - t_k} & \frac{F_2(t_k, T_m)}{T_m - t_k} \end{bmatrix}}_C \begin{bmatrix} \mathcal{X}_1(k) \\ \mathcal{X}_2(k) \end{bmatrix} + \begin{bmatrix} -\frac{E_1(t_k, T_1)}{T_1 - t_k} \\ -\frac{E_1(t_k, T_2)}{T_2 - t_k} \\ \vdots \\ -\frac{E_1(t_k, T_m)}{T_m - t_k} \end{bmatrix} + \begin{bmatrix} \mathcal{V}_1(k) \\ \mathcal{V}_2(k) \\ \vdots \\ \mathcal{V}_m(k) \end{bmatrix},$$

where

$$E(t, T) = \sum_{i=1}^2 \frac{(\kappa_i^2(\theta_i - \frac{\sigma_i \lambda_i}{\kappa_i}) - \frac{\sigma_i^2}{2})(F_i(t, T) - (T - t))}{\kappa_i^2} - \frac{\sigma_i^2 F_i^2(t, T)}{4\kappa_i},$$

and

$$F_i(t, T) = \frac{1}{\kappa_i} (1 - e^{-\kappa_i(T-t)}), \quad i = 1, 2.$$

Here, κ_i , σ_i , λ_i and θ_i , $i = 1, 2$ are model parameters. For the example from [19] considered here, $m = 6$. $\mathcal{V}_i(k) \sim N(0, H)$ are noise variables and $H = \text{diag}(h_1^2, h_2^2, \dots, h_m^2)$, where h_i are positive constants. In practice, equation (10) is the two factor short rate model, while each $\mathcal{Y}(t_k, T_i)$ denotes the yield at time t_k for maturity T_i . The parameters used in our simulation are the same as those estimated from real data in [19]. These are listed in Table I; see [19] for the exact details of the data set, parameter estimation procedure etc. As the parameters are obtained from data, one has an estimate of the standard error in each parameter. We consider a small, zero mean, normally distributed perturbation in the nominal values of κ_1 and κ_2 , with a standard deviation of $\beta = 5\%$ of nominal value of each of the two parameters. Matrices G_1 and G_2 are then computed to reflect the element-wise uncertainties introduced in matrices A and C respectively, using Monte Carlo simulation. This experiment is then repeated for a changed standard deviation of $\beta = 10\%$ and $\beta = 15\%$ of the nominal values for both the parameters.

The initial conditions as in [19] are used: $\mathcal{X}(0) = [0.015 \quad 0.025]^\top$, $\hat{\mathcal{X}}(0) = [0.02 \quad 0.02]^\top$ and $P(0) = 5 \times 10^{-3} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Our goal in this section is to see whether a small uncertainty or a small random perturbation in the parameter values has an impact on the filtering performance. We will compare the KF (which ignores the parameter uncertainty) with the filter proposed in this paper (called MKF in the tables in this section) which encapsulates the parameter uncertainty in terms of multiplicative noise, via G_1 and G_2 matrices. In

TABLE I
ESTIMATED PARAMETERS FOR VASICEK MODEL

Parameters	κ_1	θ_1	σ_1	λ_1	κ_2
Values	0.7030	0.0056	0.0321	-0.4591	0.0255
Parameters	θ_2	σ_2	λ_2	h_1	h_2
Values	0.0035	0.0142	-0.2652	0.0009	0.0012
Parameters	h_3	h_4	h_5	h_6	
Values	0.0013	0.0007	0.0009	0.0010	

order to compare the performance of the estimators, we use the root mean square error (RMSE) criterion. Consider 100 independent simulations, each with 200 data points. Denoting $\mathcal{X}^{(s)}(k)$, $k = 1, \dots, 200$ as the s^{th} set of true values of the state, and $\hat{\mathcal{X}}^{(s)}(k|k)$ as the filtered state estimate at time k for the s^{th} simulation run, the RMSE of the filter for each of the algorithms is calculated by

$$RMSE_{E_i}(s) = \sqrt{\frac{1}{200} \sum_{k=1}^{200} (\mathcal{X}_i^{(s)}(k) - \hat{\mathcal{X}}_i^{(s)}(k|k))^2},$$

$$i = 1, 2, \quad s = 1, \dots, 100.$$

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

$$AvRMSE_i = \frac{1}{100} \sum_{s=1}^{100} RMSE_{E_i}(s), \quad i = 1, 2,$$

The results for the three different levels of perturbations, $\beta = 5\%$, $\beta = 10\%$ and $\beta = 15\%$, are summarised in the table II. Recall that each of the parameters κ_1, κ_2 are perturbed by normally distributed random noise with zero mean and a standard deviation equal to β times their respective nominal values, to generate G_1, G_2 matrices in the state space equations using Monte Carlo simulation. As can be seen, the modified filter, *i.e.* the MKF has a significantly smaller AvRMSE than the KF, for both the states and for all the three levels of parameter perturbations. The RMSE values of KF are very high and are equal within four decimal places, for all three perturbation levels.

TABLE II
COMPARISON OF $AvRMSE_{E_1}$ AND $AvRMSE_{E_2}$ FOR KF AND MKF FOR DIFFERENT LEVELS OF PERTURBATIONS β

Level of perturbations		$\beta = 0.05$	$\beta = 0.10$	$\beta = 0.15$
$AvRMSE_{E_1}$	KF	1.3705	1.3705	1.3705
	MKF	0.1075	0.1469	0.1544
$AvRMSE_{E_2}$	KF	1.0811	1.0811	1.0811
	MKF	0.2368	0.2725	0.2785

B. Example 2

As another example with parameters estimated from real life data, consider a discrete-time system (1)-(2) with the following parameter specification:

$$A = \rho \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ \sin(\lambda) & \cos(\lambda) \end{bmatrix}, \quad C = [1 \quad 1],$$

$U_w^2 = 214$ and $U_v^2 = 1593$. ρ and λ are random variables with means $\hat{\rho} = 0.4$ and $\hat{\lambda} = 0.41$. The dynamics and the chosen parameter values (with $G_1 = 0$ and $G_2 = 0$) are used in ([22], chapter 2) as a time series model of rainfall in north-east Brazil. As the previous subsection, we allocate normally distributed random perturbations to λ and ρ with zero mean and β times the nominal values as the standard deviation. Three different values of β are used as in the previous case; $\beta = 5\%$, $\beta = 10\%$ and $\beta = 15\%$. The perturbation matrix G_1 is computed in each case via Monte Carlo simulation. G_2 in this case is zero since there is no uncertainty in $C = [1 \ 1]$. The initial conditions are:

$$\mathcal{X}(0) = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top, \quad \hat{\mathcal{X}}(0) = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top \quad \text{and} \\ P(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Similar to the previous subsection, the RMSE of the filter for each of the two algorithms is calculated by

$$RMSE(s) = \sqrt{\frac{1}{200} \sum_{k=1}^{200} \left(\sum_{i=1}^2 \mathcal{X}_i^{(s)}(k) - \sum_{i=1}^2 \hat{\mathcal{X}}_i^{(s)}(k|k) \right)^2}, \\ s = 1, \dots, 100,$$

Note that the error in this case is the difference between the ‘true rainfall’ generated by the model and the one step ahead rainfall prediction by the filter. The average of RMSE over 100 simulations is given by

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

Table III summarises the results of this numerical experiments. As can be seen, the proposed filtering algorithm clearly outperforms the KF, which ignores the parameter uncertainties. This is in keeping with the theoretical results and is consistent with our previous example.

TABLE III
COMPARISON OF $AvRMSE$ FOR KF AND MKF FOR DIFFERENT LEVELS OF PERTURBATIONS β

Level of perturbations		$\beta = 0.05$	$\beta = 0.10$	$\beta = 0.15$
$AvRMSE$	KF	6.3097	6.3697	6.5014
	MKF	5.0106	5.4468	5.8086

IV. CONCLUSION

In this paper, the optimal linear minimum variance filter is discussed for a class of 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 [13]. We have also provided a new interpretation of filtering under multiplicative noise in terms of filtering under parameter perturbations in an additive noise model. The results of this paper were applied to two different real data experiments for linear systems with additive-multiplicative noises. Our numerical experiments indicate that the proposed filtering algorithm can be used to improve the filtering performance (as measured by the

estimation error variance), when there is uncertainty in the estimated model parameters. We are currently investigating two different research directions: firstly, we are looking at an application of this algorithm for volatility estimation from option prices. As the state space models involved in this application are nonlinear, this would also bring in our second direction of research, which would involve exploring the use of this filtering algorithm to generate a prior for a sequential Monte Carlo filter.

APPENDIX

The filtering estimates of the state covariance is obtained by combining the equations (1)-(6) 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 proof below is a straightforward modification of a similar proof in [13] and reproduced here for the sake of completeness. The state covariance matrix at time $k + 1$ can be written as

$$\begin{aligned} \bar{P}(k+1|k+1) &= \mathbb{E}[(\mathcal{X}(k+1) - \hat{\mathcal{X}}(k+1|k+1))(\star)^\top] \\ &= \mathbb{E}[(A\mathcal{X}(k) + B + U_w\mathcal{W}(k) + G_1 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_1(k) \\ &\quad - (A\hat{\mathcal{X}}(k|k) + B + \bar{K}(k+1)(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^\top)] \\ &= \mathbb{E}[A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(\star)^\top] + U_w U_w^\top + \\ &\quad G_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top + \\ &\quad \bar{K}(k+1)(\mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^\top])\bar{K}(k+1)^\top - \\ &\quad \bar{K}(k+1)(\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + \\ &\quad G_1 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_1(k))(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))^\top] + \\ &\quad \mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) \\ &\quad + U_w\mathcal{W}(k) + G_1 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_1(k))^\top]) \end{aligned} \quad (11)$$

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

$$\begin{aligned} \mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(\star)^\top] &= CA\bar{P}(k|k)A^\top C^\top + \\ &\quad CU_w U_w^\top C^\top + CG_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top C^\top \\ &\quad + U_v U_v^\top + G_2(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_2^\top. \end{aligned} \quad (12)$$

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

$$\begin{aligned} \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + G_1 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_1(k)) \\ (\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))^\top] \\ &= \mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w\mathcal{W}(k) + \\ &\quad G_1 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_1(k))(CA\mathcal{X}(k) + CU_w\mathcal{W}(k) + \\ &\quad CG_1 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_1(k) + U_v\mathcal{V}(k) + \\ &\quad G_2 \text{diag}(\mathcal{X}_j(k))\mathcal{S}_2 - CA\hat{\mathcal{X}}(k|k))^\top] \\ &= C\mathbb{E}[(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k))^\top A^\top)] + \\ &\quad CU_w U_w^\top + CG_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top \\ &= A\bar{P}(k|k)A^\top C^\top + U_w U_w^\top C^\top + \\ &\quad G_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top C^\top. \end{aligned} \quad (13)$$

Further,

$$\begin{aligned} & \mathbb{E}[(\mathcal{Y}(k+1) - \hat{\mathcal{Y}}(k+1|k))(A(\mathcal{X}(k) - \hat{\mathcal{X}}(k|k)) + U_w \mathcal{W}(k) \\ & + G_1 \text{diag}(\mathcal{X}_j(k)) \mathcal{S}_1(k))^\top] \\ & = CA\bar{P}(k|k)A^\top + CU_w U_w^\top + \\ & CG_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top. \end{aligned} \quad (14)$$

Substituting (12), (13) and (14) in (11), we have

$$\begin{aligned} \bar{P}(k+1|k+1) & = A\bar{P}(k|k)A^\top + U_w U_w^\top + \\ & G_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top \\ & + \bar{K}(k+1)(CA\bar{P}(k|k)A^\top C^\top + CU_w U_w^\top C^\top + \\ & CG_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top C^\top + U_v U_v^\top + \\ & G_2(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_2^\top) \bar{K}(k+1)^\top \\ & - 2\bar{K}(k+1)(A\bar{P}(k|k)A^\top C^\top + U_w U_w^\top C^\top + \\ & G_1(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_1^\top C^\top) \\ & = \bar{P}(k+1|k) + \bar{K}(k+1)(C\bar{P}(k+1|k)C^\top + U_v U_v^\top + \\ & G_2(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_2^\top) \\ & \bar{K}(k+1)^\top - 2\bar{K}(k+1)\bar{P}(k+1|k)C^\top, \end{aligned} \quad (15)$$

where $\bar{P}(k+1|k)$ is as defined in (9). To find the value of $\bar{K}(k+1)$ that minimizes the trace of the covariance $\bar{P}(k+1|k+1)$ we differentiate the trace of the above expression with respect to matrix $\bar{K}(k+1)$ and set the derivative to zero.

$$\begin{aligned} \frac{\partial \text{tr} \bar{P}(k+1|k+1)}{\partial \bar{K}(k+1)} & = -2\bar{P}(k+1|k)C^\top + 2\bar{K}(k+1) \\ & [C\bar{P}(k+1|k)C^\top + U_v U_v^\top + \\ & G_2(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_2^\top]. \end{aligned} \quad (16)$$

Setting $\frac{\partial \text{tr} \bar{P}(k+1|k+1)}{\partial \bar{K}(k+1)} = 0$ leads to

$$\begin{aligned} \bar{K}(k+1) & = \bar{P}(k+1|k)C^\top [C\bar{P}(k+1|k)C^\top + U_v U_v^\top + \\ & G_2(\text{diag}(\bar{P}_{jj}(k|k) + (\hat{\mathcal{X}}_j(k|k))^2))G_2^\top]^{-1}, \end{aligned} \quad (17)$$

which is the required expression.

REFERENCES

- [1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of AMSE (Journal of Basic Engineering)*, vol. 82 D, no. 1, pp. 35–45, 1960.
- [2] S. S. Haykin, *Kalman filtering and neural networks*. Wiley, 2001.
- [3] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *IEEE*, vol. 92, pp. 401–422, 2004.
- [4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, 2002.
- [5] E. Gershon, U. Shaked, and I. Yaesh, " H_∞ -control and filtering of discrete-time stochastic systems with multiplicative noise," *Automatica*, vol. 37, no. 3, pp. 409–417, 2001.
- [6] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*. Springer, 1992, vol. 23.
- [7] K. Jordan, "A minimal intervention principle for coordinated movement," in *Advances in Neural Information Processing Systems 15: Proceedings of the 2002 Conference*, vol. 15. The MIT Press Cambridge, MA, 2003, p. 27.
- [8] R. R. Mohler, *Bilinear control processes: with applications to engineering, ecology and medicine*. Academic Press, Inc., 1973.

- [9] W. Li, E. Todorov, and 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.
- [10] F. Carravetta, A. Germani, and N. Raimondi, "Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise," *Automatic Control, IEEE Transactions on*, vol. 42, no. 8, pp. 1106–1126, 1997.
- [11] O. L. Costa and G. R. Benites, "Linear minimum mean square filter for discrete-time linear systems with markov jumps and multiplicative noises," *Automatica*, vol. 47, no. 3, pp. 466–476, 2011.
- [12] F. Yang, Z. Wang, and Y. Hung, "Robust Kalman filtering for discrete time-varying uncertain systems with multiplicative noises," *Automatic Control, IEEE Transactions on*, vol. 47, no. 7, pp. 1179–1183, 2002.
- [13] K. Ponomareva and P. Date, "An exact minimum variance filter for a class of discrete time systems with random parameter perturbations," *Applied Mathematical Modelling*, vol. 38, no. 9, pp. 2422–2434, 2014.
- [14] B. Chow and W. P. Birkemeier, "A new structure of recursive estimator," *Automatic Control, IEEE Transactions on*, vol. 34, no. 5, pp. 586–572, 1989.
- [15] B. Chow and W. Birkemeier, "A new recursive filter for systems with multiplicative noise," *Information Theory, IEEE Transactions on*, vol. 36, no. 6, pp. 1430–1435, 1990.
- [16] J. Jimenez and T. Ozaki, "Linear estimation of continuous-discrete linear state space models with multiplicative noise," *Systems & control letters*, vol. 47, no. 2, pp. 91–101, 2002.
- [17] J. Ishihara, M. Terra, and J. Cerri, "Optimal robust filtering for systems subject to uncertainties," *Automatica*, vol. 52, pp. 111–117, 2015.
- [18] O. Vasicek, "An equilibrium characterization of the term structure," *Journal of financial economics*, vol. 5, no. 2, pp. 177–188, 1977.
- [19] P. Date and R. Bustreo, "Value-at-risk for fixed-income portfolios: a Kalman filtering approach," *IMA Journal of Management Mathematics*, p. dpv016, 2016.
- [20] P. Date and C. Wang, "Linear Gaussian affine term structure models with unobservable factors: Calibration and yield forecasting," *European Journal of Operational Research*, vol. 195, no. 1, pp. 156–166, 2009.
- [21] S. H. Babbs and K. B. Nowman, "Kalman filtering of generalized vasicek term structure models," *Journal of Financial and Quantitative Analysis*, vol. 34, no. 01, pp. 115–130, 1999.
- [22] A. C. Harvey, *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.