On a New Low-Rank Kalman-Bucy Filter and its Convergence Property

Shuto Yamada\textsuperscript{1} and Kentaro Ohki\textsuperscript{1}

\textsuperscript{1}Graduate School of Informatics, Kyoto University, Kyoto, Japan
(Tel: +81-75-753-4754; E-mail: ohki@i.kyoto-u.ac.jp)

Abstract

In this paper, we propose a new low-rank Kalman-Bucy filter. We approximate the Riccati equation associated with the Kalman-Bucy filter by low-rank matrices and discuss its convergence property. Furthermore, a condition under which the proposed low rank Kalman-Bucy filter becomes stable is derived. The performance between the proposed and previous approximations are shown numerically.

1 Introduction

The Kalman/Kalman-Bucy filter has been used to estimate the system state from the sensor measurements in various fields, such as space engineering, robotics and weather forecast. However, it is difficult to implement the filter if the dimension of the state becomes high because of its computational cost \cite{1}. In order to deal with this problem, several reduction methods have been proposed in previous works using model reduction \cite{2, 3, 4}, optimization approach \cite{5} and approximation of the Kalman-Bucy filter with low-rank matrices \cite{6}. The theory of model reduction, for instance, are mature methods, but basically it requires stable linear time-invariant systems. Another method \cite{6} based on differential geometry gives an approximate method of the Riccati differential equation for general linear dynamical systems. Although it is applicable to all linear dynamical systems, the approximation is too simple and the existence of steady-state approximated filters is not provided in their paper \cite{6}. Therefore, in this paper, we propose a new low-rank Kalman-Bucy filter for general stochastic linear dynamical systems driven by white Gaussian noises and show some properties theoretically and numerically. The new low-rank Kalman-Bucy filter is similar to that of \cite{6}. The different point is that part of a statistical property of the system noise is used in our proposed method. The resultant filter includes more information on the original system’s statistics, and the accuracy is numerically evaluated in some examples.

The contributions of this paper are summarized as follows:

1. A new low-rank Kalman-Bucy filter for linear stochastic systems is proposed.

2. Some convergence and stability properties of the proposed method are derived for stochastic linear time-invariant systems.

The rest of this paper is organized as follows. In Section 2, we show the system considered in this paper and a property of the Kalman-Bucy filter. We propose a new method to approximate the Kalman-Bucy filter by low-rank matrices and show some properties including convergence and stability in Section 3. We verify the convergence property and compare the performance of the proposed low-rank Kalman-Bucy filter with other methods numerically in Section 4. This paper is concluded in Section 5. The proof of the all theoretical results are given in Appendix.

Notation

Let $\mathbb{R}$ and $\mathbb{C}$ be the set of real and complex numbers, respectively. $X^\top$ is the transpose of the real matrix $X$ and $X^*$ denotes the adjoint of the complex matrix $X$. $I_n$ is the identity matrix of size $n$. For complex numbers $a, b \in \mathbb{C}$, let $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ be their real and imaginary part, respectively, and $a \leq b$ means $\text{Re}(a) \leq \text{Re}(b)$. $P_+(n)$ is the set of symmetric positive definite $n \times n$ matrices, $S_+(r, n)$ is the set of symmetric positive semidefinite $n \times n$ matrices of rank $r$ and $\text{St}(r, n)$ is the Stiefel manifold, i.e., the set of $n \times n$ matrices with orthonormal columns: $U^\top U = I_r$.

2 Kalman-Bucy filter

As mentioned above, our method can be applied to general linear dynamical system, but in this paper, we only focus on linear time-invariant systems in order to discuss the stability of the approximated filter later. Let us consider the following continuous-time stochastic linear time-invariant system

$$
\frac{d}{dt} x(t) = Ax(t) + Gw(t),
$$

$$
y(t) = Cx(t) + Hv(t),
$$

where $x(t) \in \mathbb{R}^n$ is the state, $w(t) \in \mathbb{R}^n$ is the process noise, $y(t) \in \mathbb{R}^p$ is the measurement, and $v(t) \in \mathbb{R}^p$ is the measurement noise. The random vectors $w(t)$ and $v(t)$ are mutually independent Gaussian white noises with zero mean and unit covariance matrix. The matrices $A, G, C, H$ have the appropriate dimensions and...
\[ HH^\top \] is positive definite. Note that since we only focus on linear dynamical systems with additive Gaussian white noises, the use of stochastic differential equations is not necessary. In this case, Ito-type or Stratonovich-type stochastic differential equations are the same, and the formal calculation always coincides with the rigorous ones. Then, the minimum mean square error estimation of the state \( x(t) \) given measurements up to time \( t \) is calculated by the celebrated Kalman-Bucy filter \([7]\):

\[
\frac{d}{dt} \hat{x}(t) = (A - \hat{P}(t)C^\top(HH^\top)^{-1}C)\hat{x}(t) + \hat{P}(t)C^\top(HH^\top)^{-1}y(t),
\]

\[
\frac{d}{dt} \hat{P}(t) = A\hat{P}(t) + \hat{P}(t)A^\top + GG^\top - \hat{P}(t)C^\top(HH^\top)^{-1}C\hat{P}(t).
\]

In particular, the following property holds under certain conditions \([8, 6]\).

**Proposition 1** If \((C, A)\) is detectable and \((A, G)\) is controllable, there exists a unique solution \( Q \in P_+(n) \) of the algebraic Riccati equation \( AQ + QA^\top + GG^\top - QC^\top(HH^\top)^{-1}CQ = 0 \) and \( A - QC^\top(HH^\top)^{-1}C \) is stable. Furthermore, \( \hat{P}(t) \) converges to this stationary solution \( Q \).

### 3 Low-rank Kalman-Bucy filter

When the dimension of the state space becomes very large, the calculation of the error covariance matrix \( \hat{P} \) on the solution of Eq. (1) becomes numerically expensive. To deal with this problem, Bonnabel and Sepulchre \([6]\) proposed a method to approximate the error covariance matrix \( \hat{P} \) by a low-rank matrix. Any low-rank matrix \( P \in S_+(r, n) \) with \( r \ll n \) can be factorized as \( P = URU^\top \) where \( U \in \text{St}(r, n); R \in P_+(r) \). Using the low-rank matrix \( P = URU^\top \in S_+(r, n) \) as the approximated error covariance matrix, the estimate of the system state is calculated as follows:

\[
\frac{d}{dt} \hat{x}(t) = (A - U(t)R(t)U(t)^\top C^\top(HH^\top)^{-1}C)\hat{x}(t) + U(t)R(t)U(t)^\top C^\top(HH^\top)^{-1}y(t),
\]

However, the solution of (1) does not belong to \( S_+(r, n) \) because of the process noise covariance matrix \( GG^\top \). Therefore, Bonnabel and Sepulchre \([6]\) proposed to approximate (1) by replacing \( GG^\top \) with \( \mu UU^\top \) where \( \mu \) is the lowest eigenvalue of \( GG^\top \). This provides a low-rank Riccati equation composed of the Oja flow \([9]\) and low-dimensional Riccati equation. Unfortunately, their proposed method is not a good approximation of the original Riccati equation due to the replaced noise term. Moreover, to show the convergence property of their proposed low-rank Kalman-Bucy filter, \( A = A^\top \) is essentially required in \([6]\).

In this work, we propose a new method to approximate (1) including more information about the process noise covariance matrix. Furthermore, to analyze the stability of the proposed low-rank Kalman filter, we show some properties of the converged low-rank Kalman-Bucy filter. The proofs of all following results are given in Appendix.

Replacing \( GG^\top \) with \( U(t)U(t)^\top GG^\top U(t)U(t)^\top \), the equation (1) becomes

\[
\frac{d}{dt} \hat{P}(t) = \frac{d}{dt}\{U(t)R(t)U(t)^\top\}
= AU(t)R(t)U(t)^\top + U(t)R(t)U(t)^\top A^\top
+ U(t)U(t)^\top GG^\top U(t)U(t)^\top
- U(t)R(t)U(t)^\top C^\top(HH^\top)^{-1}CU(t)R(t)U(t)^\top,
\]

where

\[
\frac{d}{dt} U(t) = (I_n - U(t)U(t)^\top)AU(t),
\]

and \( A_{U(t)} := U(t)^\top AU(t), C_{U(t)} := CU(t) \) and \( G_{U(t)} := U(t)^\top G \). The solution of (3) belongs to \( \text{St}(r, n) \); for all \( t \geq 0 \), \( U(t)^\top U(t) = I_r \), but \( U(t)^\top U(t) \neq I_n \) if \( n \neq r \). Note that this decomposition also holds for linear time-varying systems. The equation (3) is known as the Oja flow \([9]\) and the following property is shown in \([6]\).

**Proposition 2** Suppose that \( A \) is symmetric. Let \( \lambda_1 \geq \ldots \geq \lambda_n \) be the eigenvalues of \( A \) and \( v_1, \ldots, v_n \) be their corresponding eigenvectors. If \( \lambda_r > \lambda_{r+1} \), the solution of the Oja flow (3) converges to an element of the equilibria set \( \mathcal{U} \) which spans the eigenspace associated with the \( r \) largest eigenvalues of \( A \).

Once \( U(t) \) has converged to \( U_r \in \mathcal{U} \), the equation (4) becomes a time-invariant Riccati equation, and using Prop. 1, \( R(t) \) converges to \( R_{U_r} \in P_+(r) \) if \( \{C_{U_r}, A_{U_r}\} \) is detectable and \( \{A_{U_r}, G_{U_r}\} \) is controllable. Furthermore, the following property holds.

**Proposition 3** Let \( U \) be an equilibrium solution of (3). If \((C, A)\) is detectable, \( \{C_U, A_U\} \) is also detectable and if \((A, G)\) is controllable, \( \{A_U, G_U\} \) is also controllable.

This implies that if the original system is detectable and controllable, then the approximated small-size system with \( \{A_{U(t)}, G_{U(t)}, C_{U(t)}\} \) becomes detectable and controllable, too. Therefore, the solution of (4) converges to \( R_{U_r} \) and \( A_{U_r} - R_{U_r}C_{U_r}(HH^\top)^{-1}C_{U_r} \) is stable. As \( U(t) \) and \( R(t) \) converge to \( U_r \) and \( R_{U_r} \), respectively, the approximated error covariance matrix converges to \( P = U_rR_{U_r}U_r^\top \). Since \( U_r \) is not unique, \( R_{U_r} \) is not unique, too. However, the converged error covariance matrix \( P \) is unique. This comes from the following proposition.
Proposition 4 For $U_1, U_2 \in \mathcal{U}$ and $R_{U_1}, R_{U_2}$, $U_1 R_{U_1} U_1^\top = U_2 R_{U_2} U_2^\top$, that is, the converged error covariance matrix $P$ is unique.

The converged low-rank Kalman-Bucy filter has the following property.

Theorem 1 Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of $A$ and $\sigma_1, \ldots, \sigma_r$ be the eigenvalues of $A_{U_r} = R_{U_r} C_{U_r}^\top (H H^\top)^{-1} C_{U_r}$. Then, the eigenvalues of $A - P C^\top (H H^\top)^{-1} C$ are $\sigma_1, \ldots, \sigma_r, \lambda_{r+1}, \ldots, \lambda_n$.

From this property, if $r$ is not less than the number of unstable eigenvalues of $A$, the proposed low-rank Kalman-Bucy filter becomes stable. Furthermore, we have the following proposition.

Proposition 5 If $A$ has $r$ unstable eigenvalues and $A - P C^\top (H H^\top)^{-1} C$ is stable, then rank$P \geq r$.

Prop. 4 implies that if we know the number of unstable eigenvalues of $A$, we can compose the lowest-rank Kalman-Bucy filter which becomes stable using our proposed method.

4 Numerical simulation

Here, we show the results of two numerical simulations to verify the convergence property of $U(t), R(t)$ and compare the performance of our proposed method with the original Kalman-Bucy filter and the method proposed in [6].

4.1 Simulation 1

First, let $n = 5, p = 3, r = 2$ and $A$ have $-0.3, -0.5, -0.7, -0.7, -1$ as eigenvalues. The step width $\Delta = 0.01$ and the simulation time $T = [0, 30.0]$. To verify the convergence property of the solution $U(t)$ of the Oja flow (3) and the solution $R(t)$ of the Riccati equation (4), we show the result of the numerical simulation. The components of the first column of $U(t)$, the second column of $U(t)$ and $R(t)$ are illustrated in Fig. 1, Fig. 2 and Fig. 3, respectively. From these results, we can see that the convergence property of $U(t)$ and $R(t)$ holds.

4.2 Simulation 2

Second, let $n = 10, p = 5, r = 4$ and $A, C, G$ satisfy the assumption in Prop. 1 and Prop. 2. The step width $\Delta = 0.01$ and the simulation time $T = [0, 50.0]$. In Fig. 4, the vertical axis implies the mean square error, i.e., the trace of the error covariance matrix and the horizontal axis denotes the time $t$. The error covariance matrix $V(t) := E[(x(t) - \tilde{x}(t))(x(t) - \tilde{x}(t))^\top]$ using our proposed method is calculated as follows:

$$
\dot{V}(t) = (A - P(t) C^\top (H H^\top)^{-1} C)V(t) + V(t)(A - P(t) C^\top (H H^\top)^{-1} C)^\top + GG^\top + P(t) C^\top (H H^\top)^{-1} CP(t).
$$

The blue solid line, the red dashed line, and the black dotted line represent the results of using the original Kalman-Bucy filter, our proposed approximation, and the approximation proposed by [6], respectively. These results imply that our proposed method works better than the method proposed in [6].

5 Conclusion

In this paper, we proposed a new method to compose a low-rank Kalman-Bucy filter and analyzed its convergence property and stability of the converged low-rank Kalman-Bucy filter. We also illustrated the results of two simulations to verify that the convergence property holds and our proposed low-rank Kalman-Bucy filter works better than the one proposed in [6]. However, the computational complexity was not analyzed and discrete time Kalman filter was not considered in this paper. The computational efficiency and the discrete-time version of our proposed low-rank approximation method will be tackled as future work.

References

[1] François Auger, Mickael Hilairet, Josep M Guerrero, Eric Monmasson, Teresa Orlowska-Kowalska, and Seiichiro Katsura. Industrial applications of the Kalman filter: A review. IEEE Transactions on Industrial Electronics, Vol. 60, No. 12, pp. 5458–5471, 2013.

[2] Goro Obinata and Brian D. O. Anderson. Model Reduction for Control System Design. Springer Science & Business Media, 2012.

[3] Athanasios C Antoulas. Approximation of Large-Scale Dynamical Systems. SIAM, 2009.

[4] Yuming Ba, Lijian Jiang, and Na Ou. A two-stage ensemble Kalman filter based on multiscale model reduction for inverse problems in time fractional diffusion-wave equations. Journal of Computational Physics, Vol. 374, pp. 300–330, 2018.

[5] Mathias Hudoba de Badyn and Mehran Mesbahi. Large-scale distributed Kalman filtering via an optimization approach. IFAC-PapersOnLine, Vol. 50, No. 1, pp. 10742–10747, 2017.

[6] Silvère Bonnabel and Rodolphe Sepulchre. The geometry of low-rank Kalman filters. In Matrix Information Geometry, pp. 53–68. Springer, 2013.

[7] Rudolph E Kalman and Richard S Bucy. New results in linear filtering and prediction theory. Journal of Basic Engineering, vol. 83, No. 1, Vol. 83, No. 1, pp. 95–108, 1961.
Appendix A. Proof of Proposition 3

Proof Let \( \lambda \) be an unstable eigenvalue of \( A \). Suppose that \((C_U, A_U)\) is not detectable, there exists \( \nu \neq 0 \) such that \( U^\top A U \nu = \lambda \nu \) and \( C U \nu = 0 \). From \( U^\top A U \nu = \lambda \nu \) and \( U U^\top A U = A U \), \( A U \nu = \lambda U \nu \) and \( U \nu \neq 0 \) hold. This is contradictory to the assumption that \((C, A)\) is detectable. Therefore, \((C_U, A_U)\) is detectable. The controllability of \((A_U, G_U)\) can be proved in the same way. \( \square \)

Appendix B. Proof of Proposition 4

Proof If \( U_1 \) is an equilibrium of (3), \((U_1 O)^\top (U_1 O) = I_r \) and 

\[
(I_n - (U_1 O)(U_1 O)^\top)AU_1 O \\
= (I_n - U_1 U_1^\top)AU_1 O = 0
\]

holds. Therefore, using an orthogonal matrix \( O \), \( U_2 \) can be expressed as \( U_2 = U_1 O \). The solution of the following

Fig. 3: The components of \( R(t) \)

Fig. 4: Mean Square Error
Thus, $\sigma_1, \ldots, \sigma_r$ are the eigenvalues of $A - PC^T(\mathbb{H}\mathbb{H}^T)^{-1}C$.

Since an element of the equilibria set $\mathcal{U}$ of the Oja flow (3) spans the eigenspace associated with the $r$ largest eigenvalues of $A$, $U_r \in \mathcal{U}$ can be represented as follows: $U_r = V_rO$ where $V_r := (v_1, \ldots, v_r)$ and $O$ is an orthogonal matrix.

From rank$(I_n - U_rU_r^T) = n - r$, $(I_n - U_rU_r^T)V_r = 0$ and $(I_n - U_rU_r^T)P = (I_n - U_rU_r^T)U_rR_{U_r}U_r = 0$, it follows that $(I_n - U_rU_r^T)v_j \neq 0$ ($j = r + 1, \ldots, n$) and

$$\begin{align*}
v_j^T(I_n - U_rU_r^T)(A - PC^T(\mathbb{H}\mathbb{H}^T)^{-1}C - \lambda_jI_n) & \\
& \times (I_n - U_rU_r^T)v_j
\end{align*}$$

Therefore, $\lambda_{r+1}, \ldots, \lambda_n$ are the eigenvalues of $A - PC^T(\mathbb{H}\mathbb{H}^T)^{-1}C$.

\hfill \Box

**Appendix D. Proof of Proposition 5**

**Proof** Let $\lambda_1, \ldots, \lambda_r$ and $v_1, \ldots, v_r$ be the unstable eigenvalues of $A$ and their corresponding eigenvectors.

From $A - PC^T(\mathbb{H}\mathbb{H}^T)^{-1}C$ is stable, for $j = 1, \ldots, r$, it follows that

$$\begin{align*}
v_j^T(A - PC^T(\mathbb{H}\mathbb{H}^T)^{-1}C)v_j & < 0, \\
\lambda_j - v_j^TPC^T(\mathbb{H}\mathbb{H}^T)^{-1}Cv_j & < 0.
\end{align*}$$

Considering $\lambda_j \geq 0$, $PC^T(\mathbb{H}\mathbb{H}^T)^{-1}Cv_j \neq 0$ and $C^T(\mathbb{H}\mathbb{H}^T)^{-1}Cv_1, \ldots, C^T(\mathbb{H}\mathbb{H}^T)^{-1}Cv_r$ are linearly independent because $v_1, \ldots, v_r$ are linearly independent. Thus, the rank of $P$ is not less than $r$.

\hfill \Box