Fixed Interval Optimal Estimation for Linear Discrete-Time Markovian Jump Systems by Maximum Likelihood Approach

Gou Nakura

E-mail: gg9925_fiesta@ybb.ne.jp

Abstract

In this paper we study the optimal state estimation problems for a class of linear discrete-time Markovian jump systems. We adopt maximum likelihood (ML) approach and stochastic variational calculus method to derive forms of dynamic estimators on the fixed time interval. The necessary condition for the solvability of the optimal state estimation problems are given by the coupled Riccati difference equations with initial conditions.

Key Words: Markovian jump systems; Optimal estimation; Maximum likelihood; Filtering; Smoothing

1 Introduction

Markovian jump systems ([4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 17, 18, 25, 26, 27]) have abrupt random mode changes in their dynamics. The mode changes follow Markov processes. Such systems may be found in the area of power systems, manufacturing systems, communications, aerospace systems, financial engineering and so on. Such systems are classified into continuous-time ([5, 8, 10, 11, 15, 17, 25, 26]) and discrete-time ([4, 6, 7, 9, 14, 18, 27]) systems. The optimal and $H_\infty$ control theory has presented for each of these systems respectively ([7, 9, 15, 25]). With regard to the estimation for the Markovian jump systems, the LMMSE filtering theory ([6, 7, 11]), the approximately optimal smoothing theory ([4, 14]) and the $H_\infty$ filtering theory ([26, 27]) in terms of LMIs have been presented.

Recently the author has presented the preview (noncausal) optimal and $H_\infty$ tracking theories for continuous- and discrete-time Markovian jump systems ([17, 18, 20, 21]). He has also presented the direct derivation method of fixed-preview and noncausal compensator dynamics by stochastic variational calculus ([19, 23]). From the point of view of target tracking, tracking control and state estimation are equivalent ([22]), and so noncausal optimal and $H_\infty$ state estimation theories must be able to be constructed by stochastic variational calculus. Hence the author has presented the optimal estimation theory for the continuous-time Markovian jump systems by stochastic variational approach ([24]). However the optimal estimation problems for discrete-time Markovian jump systems have not been fully investigated.

In this paper we study the optimal state estimation problems for a class of linear discrete-time Markovian jump systems. In [27], the $H_\infty$ filtering theory for linear discrete-time Markovian jump systems has been already presented in terms of LMIs. However the direct derivation method of coupled Riccati difference equations and the relationship between filtering and smoothing theory have not been fully investigated for the systems. Also note that there exist remarkable differences between continuous-time and discrete-time Markovian jump systems. For example, on the processes of the stochastic variational calculus, we need to consider the effects of coupling terms for the continuous-time Markovian jump systems, while we do not need to consider them but need to consider the $\sigma$-fields generated by the pair of system state and mode for the discrete-time Markovian jump systems. In this paper by considering the optimal state estimation problems for the linear discrete-time jump systems we give a basis of state estimation theories and the relation filtering and smoothing for such systems. We need to clarify how the coupled Riccati difference equations and the forms of estimators can be derived. We adopt maximum likelihood (ML) approach and stochastic variational calculus method to derive an estimate and a form of a dynamic estimator on the fixed time interval. Different from the cases of ML approach to optimal state estimations for single mode systems, Markov property of system state itself does not hold ([7]). Therefore we consider a pair of system state and mode state as a joint process, for which Markov property holds. Finally we consider numerical examples and verify the effectiveness of the state estimation theory presented in this paper.

Notations: Throughout this paper the superscript ”$^\prime$” stands for the matrix transposition, $\| \cdot \|$ denotes the Euclidean vector norm and $\| \cdot \|_P$ also denotes the weighted norm $\Psi_1$. $O$ denotes the matrix with all zero components. $P(\cdot \mid \cdot)$ and $E(\cdot \mid \cdot)$ denote the conditional probability and expectation.

2 Problem formulation

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and, on this space, consider the following linear discrete-time time-
Discrete-Time Markovian Jump Systems

We adopt maximum likelihood (ML) approach in value \( \hat{\mathbf{x}}(k) \) of a likelihood function defined below. In order to solve this minimization problem we calculate the stochastic first variation of the likelihood function under the dynamics constraint.

\section{Design of Optimal Estimators}

\subsection{Maximum Likelihood Approach}

The fixed-interval smoothed estimate of the state \( x(k) \) denoted by \( \hat{x}(k|N) \) is Bayesian maximum a posteriori (MAP) estimate that maximizes the posterior smoothing density \( P(x(k), m(k)|Y_N) \).

Define

\[
X^N := \{ x_0, x(1), \ldots, x(N) \}
\]

\[
M^N := \{ m_0, m(1), \ldots, m(N) \}
\]

and

\[
Y^N := \{ y(0), y(1), \ldots, y(N) \}.
\]

By Bayes’ rule

\[
P(X^N, M^N|Y^N) = \frac{P(Y^N|X^N, M^N)P(X^N, M^N)}{P(Y^N)}
\]

By the Markov property

\[
P(X^N, M^N) = P(x_0, m_0)\Pi_{l=0}^{N-1}P(x(l+1), m(l+1)|x(l), m(l)).
\]

From the assumption \( x_0 \sim \mathcal{N}(\bar{x}_0, \Sigma_0) \)

\[
P(x(0)) = (2\pi)^{-n/2}|\Sigma_0|^{-1/2} \times \exp \left\{ -\frac{1}{2} (x_0 - \bar{x}_0)^2 \Sigma_0^{-1} \right\}.
\]

By the Gaussian property of \( w_d(l) \),

\[
P(x(l+1), m(l+1)|x(l), m(l)) = (2\pi)^{-n/2}|\tilde{Q}_{d,m}(l)|^{-1/2} \exp \left\{ -\frac{1}{2} (x(l+1) - \tilde{A}_{d,m}(l)x(l)) \right\}
\]

where \( \tilde{Q}_{d,m}(l) = G_{d,m}(k)Q_{d,m}(k)G_{d,m}^T(l) \).

Taking the product from \( l = 0 \) to \( l = N - 1 \),

\[
\frac{1}{c_1} \Pi_{l=0}^{N-1}P(x(l+1), m(l+1)|x(l), m(l))
\]

\[
= \exp \left\{ \sum_{l=0}^{N-1} -\frac{1}{2} w_d(l)^2 \tilde{Q}_{d,m}(l) \right\}
\]

with

\[
x(k+1) = A_{d,m}(k)x(k) + G_{d,m}(k)w_d(k)
\]
where \( c_3 \) is the normalizing constant. From the independence of \( v(k) \), using the observation

\[
y(k) = H_{d,m}(k)x(k) + v(k),
\]

\[
P(Y^N|X^M, M^N) = \Pi_{s=0}^N P(y(l)|x(l), m(l)) = (2\pi)^{-\frac{N+1}{2}} \Pi_{s=0}^N |R_{d,m}(l)|^{-1/2} \times \exp \left\{ -\frac{1}{2} \sum_{l=0}^{N} \|y(l) - H_{d,m}(l)x(l)\|^2_{R_{d,m}(l)} \right\}
\]

Now we obtain

\[
\frac{1}{c_2} P(X^N, M^N|Y^N) = c_3 \exp(J_{DN})
\]

where \( c_2 \) and \( c_3 \) are the normalizing constants, and \( c_3 \) does not depend on \( X^N \) and \( M^N \). Now we define

\[
J_{DN} := \mathbb{E} \left\{ \frac{1}{2} \|x_0 - \hat{x}_0\|^2_{\Sigma_0^{-1}} + \frac{1}{2} \sum_{k=0}^{N} \|H_{d,m}(k)x(k) - y(k)\|^2_{R_{d,m}(k)} + \frac{1}{2} \sum_{k=0}^{N-1} \|w_d(k)\|^2_{Q_{d,m}(k)} \right\}
\]

We consider this likelihood function \( J_{DN} \). In order to obtain the maximum a priori (MAP) estimate \( \hat{x}(k|N) \) maximizing the probability density function for the given observed information \( Y_N \), we consider the problem of seeking the \( \hat{w}_d(k) = \hat{w}_d(k|N) \) minimizing the performance index \( J_{DN} \) under the constraint of the dynamics of the system (1). In order to solve the estimation problems for the system (1) we introduce the Lagrange multiplier \( \lambda_{d,m}(\cdot) \) depending on the mode transition and define the following functional under the dynamics constraint:

\[
J_{DN} := \mathbb{E} \left\{ \frac{1}{2} \|x_0 - \hat{x}_0\|^2_{\Sigma_0^{-1}} + \frac{1}{2} \sum_{k=0}^{N} \|H_{d,m}(k)x(k) - y(k)\|^2_{R_{d,m}(k)} + \frac{1}{2} \sum_{k=0}^{N-1} \|w_d(k)\|^2_{Q_{d,m}(k)} \right\}
\]

\[
\sum_{k=0}^{N-1} \lambda_{d,m}(k) (l+1) \left\{ x(l+1) - A_{d,m}(l)x(l) - G_{d,m}(l)w_d(l) \right\}
\]

where \( \lambda_{d,m}(\cdot) \) is a Lagrange multiplier.

Now the optimal estimation problems for the system (1) can be reduced to the problem to decide \( w_d(k) \) maximizing \( J_{DN} \).

**Remark 3.1** As described in the introduction, Markov property of the system state process \( \{x(\cdot)\} \) does not hold but Markov property of the joint process \( \{x(\cdot), m(\cdot)\} \) holds. Therefore we regard the pair of the system state \( x(\cdot) \) and mode state \( m(\cdot) \) as the joint process and derive the likelihood function \( J_{DN} \) by the Markov property of it. This approach is different from the cases for single mode systems.

### 4 Design of Optimal Estimators

#### 4.1 Optimization for \( x_0 \) and \( w_d \)

By calculating the first variation \( \delta J_{DN} \) with regard to \( x_0 \) and \( \lambda_{d,m}(\cdot) \) (see the appendix 1) and letting \( \delta J_{DN} = 0 \), we obtain

(i) The conditions of optimality:

\[
w_d^*(k) = Q_{d,m}(k)G_{d,m}(k)\mathcal{E}_m(k)(\lambda_{d,m}^*(k+1), k)
\]

(ii) The canonical equations:

\[
\mathbb{E}\{x^*(k+1)|G_k\} = A_{d,m}(k)x^*(k) + G_{d,m}(k)w_d^*(k),
\]

\[
\lambda_{d,m}^*(k) = \lambda_{d,m}^*(k+1), k)
\]

\[
-H_{d,m}(k)R_{d,m}(k)(H_{d,m}(k)x^*(k) - y(k))
\]

(iii) The boundary conditions:

\[
x_0^* = \hat{x}_0 + \Sigma_0 \lambda_{d,m}(0),
\]

\[
\lambda_{d,m}(N)(N) = 0
\]

where \( G_k \) is the \( \sigma \)-field generated by the random variables \( \{x(l), m(l) : l = 0, \cdots, k\} \).

\[
\mathcal{E}_m(k)(\lambda_{d,m}^*(k+1), k) = \sum_{j=0}^{N} \pi_{d,m}(j)X_{j}(k+1) \]

\[
\lambda_{d,m}^*(k) = (\lambda_{d,m}^*(1), \cdots, \lambda_{d,m}^*(N), k)
\]

From the boundary condition (6), let

\[
x^*(k) = \hat{x}_k + X_m(k)\lambda_{d,m}(k)
\]

and then

\[
\mathbb{E}\{x^*(k+1)|G_k\} = \hat{x}_{k+1} + X_k^*(k+1, k) \mathcal{E}_m(k)(\lambda_{d,m}^*(k+1), k)
\]

where \( \mathcal{E}_m(k)(X(k+1), k) = \sum_{j=1}^{N} \pi_{d,m}(j)X_j(k+1) \)

and \( X(k) = (X_1(k), \cdots, X_N(k)) \).

Using (8) and (9), we obtain the following two equalities to vanish \( x^*(k) \) and \( \mathbb{E}\{x^*(k+1)|G_k\} \):

\[
\mathbb{E}\{\hat{x}_{k+1}|G_k\} = A_{d,m}(k)^{-1} \mathcal{E}_m(k)(\lambda_{d,m}^*(k+1), k)
\]

\[
\mathbb{E}\{\hat{x}_{k+1}|G_k\} - A_{d,m}(k)^{-1} \mathcal{E}_m(k)(\lambda_{d,m}^*(k+1), k)
\]

\[
[A_{d,m}(k)^{-1} \mathcal{E}_m(k)(\lambda_{d,m}^*(k+1), k) + A_{d,m}(k)\mathcal{X}_m(k)(\lambda_{d,m}^*(k+1), k)
\]

\[
\lambda_{d,m}^*(k) = \Psi^{-1}_m(k)\{\mathcal{A}_m(k)^{-1}\mathcal{E}_m(k)(\lambda_{d,m}^*(k+1), k)
\]

\[
+H_{d,m}(k)R_{d,m}(k)(y(k) - H_{d,m}(k)X_m(k))
\]

where

\[
\Psi_m(k) = [I_n + H_{d,m}(k)R_{d,m}(k)H_{d,m}(k)]
\]
Substituting (11) into (10), we obtain the following equality.

\[ A_{d,m}(k)X_{m}(k)\Psi^{-1}_{m}(k)A'_{d,m}(k) + G_{d,m}(k)Q_{d,m}(k)G'_{d,m}(k) - E_{m}(k)(X(k+1), k)E_{m}(k)(m_{k}(k+1), k) = E\{\hat{x}_{k+1}|G_k\} - A_{d,m}(k)\hat{x}_k - A_{d,m}(k)X_{m}(k)\Psi^{-1}_{m}(k)k \times H'_{d,m}(k)R_{d,m}(k)[y(k) - H_{d,m}(k)\hat{x}_k] \]

Since the above equality holds for arbitrary \( m_{k}(k+1) \), \( i = 1, \cdots, N^* \), we have the following two equalities.

\[ E\{\hat{x}_{k+1}|G_k\} = A_{d,m}(k)\hat{x}_k + A_{d,m}(k)X_{m}(k)\Psi^{-1}_{m}(k)k \times H'_{d,m}(k)R_{d,m}(k)[y(k) - H_{d,m}(k)\hat{x}_k] \]

\[ E_{m}(k)(X(k+1), k) = A_{d,m}(k)X_{m}(k)\Psi^{-1}_{m}(k)A'_{d,m}(k) + G_{d,m}(k)Q_{d,m}(k)G'_{d,m}(k) \]

Then we obtain the following proposition, which gives fundamental dynamics and coupled Riccati difference equations for obtaining a solution of the Optimal Estimation Problems for the system (1).

**Proposition 4.1** Consider the system (1) and the performance index (2). Assume the conditions A1 and A2. Then, if the performance index (2) is optimized with regard to \( x_0 \) and \( w_d \), i.e., \( (x_0^*, \{w_{d}^*\}) \) minimizing (2) is obtained, we obtain the dynamics (12) for deciding an optimal dynamic estimator where the positive definite matrices \( X_i, i = 1, \cdots, N^* \) are the solution of the coupled Riccati difference equations (13) with \( X_i(0) = \Sigma_0 \). These (12) and (13) give the necessary conditions for the solvability of the optimal estimation problems for (1).

Based on this proposition, we can derive the optimal filter and smoother for (1).

### 4.2 Optimal Filtering

In this case, at the current time \( k \), the output \( y(l) \), \( 0 \leq l \leq k \) is available.

Then we have the following theorem, which gives the solution of the Optimal Filtering Problem for (1). (Refer to [1, 2, 28] and so on in the single mode cases.)

**Theorem 4.1** Consider the system (1). Assume the conditions A1 and A2. The Optimal Filtering Problem for (1) is solvable if and only if there exist positive definite matrices \( X_i(k), i = 1, \cdots, N^* \), satisfying the conditions \( X_i(0) = \Sigma_0, i = 1, \cdots, N^* \), such that the coupled Riccati difference equations (13) hold over \([0, N]\). Moreover the optimal dynamic filter is given as follows:

\[
\hat{x}_{k+1|k} = A_{d,i}\hat{x}_{k|k},
\]

\[
\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_{d,i}(k)[y(k) - H_{d,i}(k)\hat{x}_{k|k-1}],
\]

\[
\hat{x}_{0|0} = \hat{x}_0,
\]

\[
K_{d,i}(k) = X_i(k)H'^{-1}_{d,i}(k)X_i(k)H'^{-1}_{d,i}(k) + R_{d,i}(k)^{-1},
\]

\[
i = 1, \cdots, N^*
\]

#### 4.3 Optimal Smoothing

In this case, at the current time \( k \), \( 0 \leq k \leq N \), all observation information \( Y_N \) is available. Then the equality (5) is reduced to

\[
\lambda_{d,m}^*(k) = A'^{t}_{d,m}(k)[A'^{t}_{d,m}(k)\lambda^*_m(k+1), k) - H'^{t}_{d,m}(k)R^{-1}_{d,m}(k) (y(k) - H_{d,m}(k)\hat{x}_k)] \]

Then we obtain the following smoothing algorithm

**Theorem 4.2** Consider the system (1). Assume the conditions A1 and A2. Then the optimal smoothing algorithm, which gives the solution of The Optimal Smoothing Problem for (1), is given as follows:

- **Optimal smoothing algorithm**
  
  *i* Obtain the smoothed estimate \( \hat{x}_{k|N} \), \( k \in [0, N] \).

- **ii** Solve the following coupled Riccati difference equations (13) with the initial conditions \( X_i(0) = \Sigma_0 \).

- **iii** Obtain the filtered estimate \( \hat{x}_{k|k} \), \( k \in [0, N] \).

**Remark 4.1** Notice that the unified ML and stochastic variational approach by which the filtering and smoothing problems is equivalent to that for the noncausal tracking control problems ([19, 23]) from the point of view of the target tracking. Also notice that the coupled Riccati difference equations (13) don’t include the mode distributions at each time, which are different from those in [7, 29].

### 5 Numerical Examples

In this section, we study numerical examples to demonstrate the effectiveness of the presented design theory.

We consider the following two mode system and assume that the system parameters are as follows:

\[
x(k+1) = A_{d,m}(k)x(k) + G_{d,m}(k)w(k),
\]

\[
x(0) = x_0, m(0) = i_0
\]

\[
y(k) = H_{d}(k)x(k) + v(k)
\]

- 50 -
where

- Mode 1:
  \[ A_{d,1} = \begin{bmatrix} -0.6 & 1 \\ 0 & 0.7 \end{bmatrix}, \quad \text{Mode 2:} \quad A_{d,2} = \begin{bmatrix} -0.4 & 0.8 \\ 0.2 & 0.5 \end{bmatrix}, \]

\[ G_d = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad H_d = [1, 0] \]

and

\[ Q_d = 1, \quad R_d = 0.1. \]

Let

\[ P_d = \begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix} \]

be a transition matrix of \( \{m_t\} \). We set \( \bar{x}_0 = \text{col}(-0.1, 0) \) and \( i_0 = 1 \). \( w_d(\cdot) \) and \( v(\cdot) \) are zero mean white noises with Gaussian distribution. The paths of \( m(k) \) are generated randomly, and the performances are compared under the same circumstance, that is, the same set of the paths so that the performances can be easily compared.

We consider the whole system (16) with mode transition rate \( P_d \) over the time interval \( k \in [0, 44] \). We design optimal filter and smoother for the system (16) according to Theorem 4.1 and Theorem 4.2. We verify the effectiveness of the filter and the smoother, and compare the estimation performances for them.

Fig. 1 shows the first component of the state affected by the Gaussian noise and the filtered estimate. Fig. 2 shows the first component of the state affected by the Gaussian noise and the smoothed estimate. Fig. 3 shows the square errors between the first components of the states and filtered values, and the states and smoothed values respectively. From these figures and calculation results it is shown that the smoother gives better estimation than the filter because of the utilization of noncausal information of the observation.

Remark 5.1 Notice that, while we need to obtain the solutions of the coupled Riccati difference equations (13) to get the filter and smoothed estimates, we do not need to compute the mode distributions at each time from the mode transition matrix \( P_d(k) \). On the algorithms in [7, 29], we need to compute the mode distribution over the whole time interval. Especially if the mode transitions are time-varying, the computation cost for the execution of the algorithms is much higher than the method presented in this paper.

6 Concluding Remarks

In this paper we have studied the optimal state estimation problems for linear discrete-time Markovian jump systems, a class of linear discrete-time hybrid systems. We have adopted maximum likelihood (ML)
approach and stochastic variational calculus method to derive the form of dynamic estimators on the fixed time interval.

We have reduced the optimal estimation problems to the optimization problems by ML approach and derived the fundamental dynamics for the optimal estimators design and coupled Riccati difference equations, which give the necessary conditions for the solvability of the optimal state estimation problems, by the stochastic variational calculi. Then we have derived the optimal filter and smoother based on these fundamental dynamics and coupled Riccati difference equations. Note that, for the Markovian jump systems, the filter dynamics are uncoupled with each other, but the smoother dynamics are coupled with each other, while the Riccati difference equations, by the solutions of which the filter and smoother gains are determined, are coupled with each other. Also note that the coupled Riccati difference equations obtained in this paper is rather natural both in the meaning that they are an extension of the Riccati equation for the single mode systems and in the meaning that they are a limit from the Riccati equations for $H_\infty$ state estimation as a disturbance index $\gamma$ to $\infty$ compared the references [7, 27, 29] and so on.

The author has already presented the optimal estimation theory for the linear continuous-time Markovian jump systems ([24]). Note that there exist remarkable differences between continuous-time and discrete-time Markovian jump systems. In fact, on the processes of the stochastic variational calculus, we need to consider the effects of coupling terms for the continuous-time Markovian jump systems, while we do not need to consider them but need to consider the $\sigma$-fields generated by the pair of system state and mode for the discrete-time Markovian jump systems.

Throughout this paper it is assumed that the modes of the system are fully observable over the whole time interval. The estimation theory in the case with inaccessible modes is a very important further research issue.

References


[17] G. Nakura: $H_\infty$ Tracking with Preview for Linear Continuous-Time Markovian Jump Systems,
Appendix 1: Calculus of Variation

In this appendix we present the calculation process of the first variation of $J_{dN}$.

For each $1 \leq k \leq N - 1$, we obtain

$$\delta J_{dN} = E\left\{\left(\|x_0 - \hat{x}_0\|_{2,1}^2 + \sum_{l=0}^{N-1} \|w_d(l)\|^2_{Q_{d,m(l)}(l)} \right.\right.$$

$$\left. + E\{\|y(l) - H_{d,m(l)}(l)x(l)\|^2_{R_{d,m(l)}(l)}\} \right\}$$

$$+ 2\gamma^2 E_{m(k)}(x_d(k + 1) + 1) \{E\{x(k + 1)\}\}G_k$$

$$- A_{d,m(k)}x(k) - G_{d,m(k)}w_d(k)\}$$

$$+ 2\gamma^2 \sum_{l=0}^{N-1} E\{x_{d,m(l+1)}(l + 1)\{x(l + 1)\}G_k$$

$$- A_{d,m(l)}x_d(l) - G_{d,m(l)}w_d(l)\}\}$$

We calculate the first variation of $J_{dN}$ with regard to $x$, $w_d$ and $\lambda_{d,m(\cdot)}$ and obtain the following equality.