Design and Stability of Moving Horizon Estimator for Markov Jump Linear Systems

Qing Sun, Cheng-Chew Lim, *Senior Member, IEEE*, Peng Shi, *Fellow, IEEE*, and Fei Liu, *Member, IEEE*

***Abstract*—This paper presents a moving horizon algorithm with mode detection for state estimation in Markov jump systems with Gaussian noise. This state estimation scheme is a combination of the maximum-likelihood algorithm and the moving horizon approach. The maximum-likelihood algorithm provides optimal estimate of the mode sequence within a moving fixed-size horizon, and the moving horizon estimation is an optimization-based solution. As a result, a mode detection-moving horizon estimator design method is proposed. Through the stochastic observability properties of the Markov jump linear systems, sufficient conditions for stability are established.**

***Index Terms*—Markov jump systems, maximum-likelihood algorithm, moving horizon approach, state estimation.**

I. Introduction

MARKOV jump linear systems (MJLS) are a class of stochastic hybrid systems with parameters governed by a finite state Markov chain. In past decades, the state estimation and control problem for the MJLS has been applied in a variety of fields including maneuvering aircraft [1], switching communication networks [2]–[5], sensor scheduling, and information fusion, see [6]–[11]. In conventional framework, the hybrid state suboptimal estimation problem can be solved by the multimodel filtering algorithm based on a bank of Bayes rule-based [12], [13] filters including the interacting multiple model (IMM) algorithm and the generalized pseudo-Bayesian algorithm, see for example, [14] for details. While suboptimal solution is adequate in some applications, finding the optimal state estimates of an MJLS in a real-time data processing environment is highly desirable, but the problem is largely open. One must overcomes the exponential complexity incurring in computing the optimal state estimates of a MJLS due to its complex dynamical properties.

One optimization-based algorithm widely used for state estimation of the hybrid systems [15]–[19] is moving horizon estimation (MHE) algorithm [20]–[23]. In the MHE algorithm, the state estimate is determined online by solving an optimization problem within a fixed-size moving horizon. Since the optimization problem is restricted within a finite dimension, the MHE algorithm is well suited for practical implementation using only limited computer memory. Originally, the idea of MHE approach was inspired by the success of optimization-based moving horizon strategy used in model predicted control [24], [25]. The MHE approach was then extended to handle nonlinearities and constraints on the state and disturbance [26]. More recently, the approach is applied to the optimal state estimation problem for hybrid systems. More specifically, the state estimation problem for hybrid systems was addressed by [21] using MHE strategy in the mixed logic dynamical systems form and piecewise affine systems. In addition, the MHE approach was used for state estimation problem of switching systems of linear and nonlinear cases [20], [22]. Therefore, it is appealing to apply the MHE approach to construct a new estimator to compute the optimal state estimate of MJLS.

In this paper, we consider the state estimation problem of a class of MJLS in the presence of Gaussian system and measurement noise, and present a new and effective MHE algorithm scheme, where the maximum-likelihood (ML) approach is introduced for detecting uncertain mode jumping processes within the MJLS. Various forms of mode detection algorithm based on the maximum likelihood procedure have been proposed [27]–[29]. Here, the ML solution takes three steps. First, calculate an optimal estimate in a *maximum a posteriori* sense by maximizing a log-likelihood function of the conditional density function of the hybrid state from a given set of noisy measurement data. Second, obtain an optimal estimate of the mode jump sequence within a moving horizon by maximizing the conditional probability of mode jump sequence with respect to the measurement data. Third, integrate the mode sequence ML-estimate into a

Manuscript received November 20, 2016; revised November 21, 2016 and September 1, 2017; accepted February 21, 2018. Date of publication March 15, 2018; date of current version February 26, 2019. This work was supported in part by the National Nature Science Foundation of China under Grant 61773131, Grant 61773183, Grant 61633016, Grant 61473182, and Grant U1509217; in part by the Australian Research Council under Grant DP170102644; and in part by the Shanghai Sail Plan for Talents Development (18YF1408100). Recommended by Associate Editor S. Yuksel. *(Corresponding author: Peng Shi.)*

Q. Sun is with the Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China, and also with the Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Institute of Automation, Jiangnan University, Wuxi 214122, China (e-mail: qingsun@live.com).

C.-C. Lim and P. Shi are with the School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, SA 5005, Australia (e-mail: cheng.lim@adelaide.edu.au; peng.shi@adelaide.edu.au).

F. Liu is with the Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Institute of Automation, Jiangnan University, Wuxi 214122, China (e-mail: fliu@jiangnan.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2018.2816102

MHE framework to obtain the finial state estimates within the fixed-size sliding horizon.

The advantage of this new MHE estimator is its fast online computational capability by having the ML removing the improper jumping sequences, and the optimal solution is available instead of solving the optimization of a bank of MHE estimators on all potential mode jumping sequences. Note that it is difficult to give sufficient conditions that guarantee the stability of the estimation error with the proposed estimation scheme in the presence of the Gaussian noise. To overcome this difficulty, an explicit bounding sequence is constructed to ensure the convergence of the mean of the norm of the estimation error when the MJLS is stochastic observable.

The main contributions of the study are threefold.

1) Combined with ML approach, a mode detection-moving horizon estimator (MD-MHE) scheme is designed for the discrete-time MJLS.
2) Three functional estimators based on MD-MHE scheme are presented, and they are called the mode detection-moving horizon estimator, the extended mode detection-moving horizon estimator (EMD-MHE), and incomplete mode detection-moving horizon estimator.
3) Based on the stochastic observability analysis for the MJLS with Gaussian noise, stability analysis of all the proposed estimators in terms of sufficient conditions to guarantee the convergence of mean square errors are provided.

The rest of this paper is organized as follows. Section II describes a model of the MJLS and discuss the formulation of the associated state estimation problem. Section III analyzes the observability of the hybrid MJLS systems in the presence of Gaussian noise. Section IV presents the estimation procedures of the MD-MHE algorithm, extended MD-MHE algorithm, and incomplete MD-MHE algorithm. Section V gives the stability analysis and proves the convergence of the estimating errors for the proposed estimators. Section VI demonstrates the performances of the proposed algorithms via computer simulations. We conclude this paper in Section VII.

Notation: Let $\mathbb{R}^n$ be the real n-dimensional Euclidean space, and $\mathbb{R}^{n\times s}$ be the space of all $(s\times n)$-dimensional real matrices. For $\Gamma\in\mathbb{R}^{n\times n}$, Γ' represents the transpose of Γ, and $\Gamma>0$ denotes Γ being real symmetric and positive definite. The eigenvalues of Γ are denoted by $\lambda_i(\Gamma)$, $i=\{1,2,\ldots,n\}$. Then, $\underline{\lambda}(\Gamma)$ $\left(\bar{\lambda}(\Gamma)\right)$ is the minimum (maximum) eigenvalue of Γ. Given a vector $z\in\mathbb{R}^n$, $\|z\|^2$ represents its Euclidean norm of z, and given $\Psi>0$, $\Psi\in\mathbb{R}^{n\times n}$, $\|z\|_\Psi\triangleq(z'\Psi z)^{\frac{1}{2}}$. Given N square matrices $\Lambda_1,\Lambda_2,\ldots,\Lambda_N$ on the space $\mathbb{R}^{n\times n}$, $\operatorname{diag}(\Lambda_1,\Lambda_2,\ldots,\Lambda_N)\in\mathbb{R}^{Nn\times Nn}$ is a block-diagonal matrix with $\Lambda_1,\Lambda_2,\ldots,\Lambda_N$ on the main diagonal and zeros outsides. A probability space is defined as $(\Omega,\mathfrak{F},\Pr)$ with the sample space Ω, σ-algebra $\mathfrak{F}$ of all subsets of Ω, and probability measure $\Pr(\cdot)$. The symbol $\Phi(x;\bar{x},P)$ denotes normal distribution with mean vector $\bar{x}$ and covariance matrix P, and its probability density is $\phi(x;\bar{x},P)$. The symbol $f(\cdot)$ denotes the probability density function, and $\mathrm{E}\{\cdot\}$ stands for the expectation operator. The chi-squared distribution is denoted by $\chi^2(\cdot)$, and the cumulative distribution function of a χ^2 random variable with n degrees of freedom is denoted by $\gamma_n(\cdot)$. The cardinality of a finite set $\mathbb{S}$ is denoted by $\operatorname{Card}(\mathbb{S})$. The symbol $\mathbb{T}$ denotes the discrete-time set $\{0,1,2,\ldots\}$.

II. Problem Formulation and Preliminaries

In the probability space $(\Omega,\mathfrak{F},\Pr)$, a class of discrete-time MJLS for $t\in\mathbb{T}$ is characterized as

$$\Sigma:\begin{cases}x_{t+1}=A_{r_t}x_t+w_t,\\ y_t=C_{r_t}x_t+v_t\end{cases}\tag{1}$$

where $x_t\in\mathbb{R}^{n_x}$ represents the unknown state vector, and $y_t\in\mathbb{R}^{n_y}$ represents the measurement vector. $w_t\in\mathbb{R}^{n_x}$ and $v_t\in\mathbb{R}^{n_y}$ are the process and measurement noise vectors with zero mean and covariance matrices Q_w and R_v.

Define a discrete-time, homogeneous, first-order Markov chain $\{r_t\}$ on a finite state space $\mathbb{M}\triangleq\{1,2,\ldots,m\}$ with transition probability matrix $\Pi=\{\pi_{ij}\}_{m\times m}$ for $i,j\in\mathbb{M}$. Here, $\pi_{ij}=\Pr\{r_{t+1}=j|r_t=i\}$, with $\pi_{ij}\geq 0$ and $\sum_{j=1}^m\pi_{ij}=1$.

The initial mode probability distribution is denoted by $\Pr(r_0=i)\triangleq p_0^i$. Here, the parameters $A_{r_t}\in\mathbb{R}^{n_x\times n_x}$ and $C_{r_t}\in\mathbb{R}^{n_y\times n_x}$ evolve with time according to the Markovian stochastic process $\{r_t\}$, and assumed to be known for each value of r_t. For brevity, let $r_t^j\triangleq\{r_t=j\}$, $A_t^j\triangleq\{A_{r_t},r_t=j\}$, and $C_t^j\triangleq\{C_{r_t},r_t=j\}$. It is easily seen that the joint process $\{x_t,r_t\}$ is Markov process, and the mode variable r_t is independent of x_0, w_t, and v_t.

Assumption 2.1: Let $P_0\in\mathbb{R}^{n_x\times n_x}$, $\mu_w\in\mathbb{R}^+$, and $\mu_v\in\mathbb{R}^+$ with $P_0>0$, $\mu_w>0$, and $\mu_v>0$.

1) The initial state distribution $x_0\sim\Phi(x_0;\bar{x}_0,P_0)$, $P_0>0$.
2) For any $t,l\in\mathbb{T}$, w_t, and v_t are Gaussian noise sequences satisfying $\mathrm{E}(w_t)=0$, $\mathrm{E}(w_t'w_l)=\mu_w\delta_{tl}$, $\mathrm{E}(v_t)=0$, and $\mathrm{E}(v_t'v_l)=\mu_v\delta_{tl}$, where $\delta_{tl}=1$, for $t=l$, and $\delta_{tl}=0$ for $t\neq l$.
3) The variables x_0, w_t, and v_t are mutually independent.

We now establish the recursive equations for Q_w, R_v, μ_w, and μ_v.

Proposition 2.1: Consider model (1). For every $t\in\mathbb{T}$

$$\mathrm{E}\{\|w_t\|^2\}=E(w_t'w_t)=\mu_w=\operatorname{tr}(Q_w)$$
$$\mathrm{E}\{\|v_t\|^2\}=E(v_t'v_t)=\mu_v=\operatorname{tr}(R_v).$$

For $s,t\in\mathbb{T}$, $\mathbf{y}_{s,t}$ denotes the measurement sequence $\{y_s',y_{s+1}',\ldots,y_t'\}'$, $\mathbf{x}_{s,t}$ denotes the state sequence $\{x_s',x_{s+1}',\ldots,x_t'\}'$, $\mathbf{r}_{s,t}$ denotes the mode sequence $\{r_s,r_{s+1},\ldots,r_t\}$, $\mathbf{w}_{s,t}$ represents $\{w_s',w_{s+1}',\ldots,w_t'\}'$, and $\mathbf{v}_{s,t}$ represents $\{v_s',v_{s+1}',\ldots,v_t'\}'$ within the time interval $[s,t]$, $s<t$. For brevity, let $\mathcal{F}(\mathbf{r}_s)\triangleq A_{r_s}$ and $\mathcal{F}(\mathbf{r}_{s,t})\triangleq A_{r_s}A_{r_{s+1}}\ldots A_{r_t}$. Then, the state sequence $\mathbf{x}_{0,t+1}$ and its measurement sequence $\mathbf{y}_{0,t}$ with the $(t+1)$-step transition map for system Σ in (1) are denoted by

$$\begin{cases}\mathbf{x}_{0,t+1}=\mathbf{F}(\mathbf{r}_{0,t})x_0+\mathbf{B}(\mathbf{r}_{0,t})\mathbf{w}_{0,t}\\ \mathbf{y}_{0,t}=\mathbf{H}(\mathbf{r}_{0,t})x_0+\mathbf{G}(\mathbf{r}_{0,t})\mathbf{w}_{0,t-1}+\mathbf{v}_{0,t}\end{cases}$$

where

$$\mathbf{F}(\mathbf{r}_{0,t}) = [I_{n_x\times n_x}, \mathcal{F}(\mathbf{r}_0), \mathcal{F}(\mathbf{r}_{0,1}), \dots, \mathcal{F}(\mathbf{r}_{0,t})]'$$

$$\mathbf{H}(\mathbf{r}_{0,t}) = [C_{r_0}, (C_{r_1}\mathcal{F}(\mathbf{r}_0)), \dots, (C_{r_t}\mathcal{F}(\mathbf{r}_{0,t-1}))]'$$

$$\mathbf{B}(\mathbf{r}_{0,t}) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ I_{n_x\times n_x} & 0 & \cdots & 0 \\ \mathcal{F}(\mathbf{r}_0) & I_{n_x\times n_x} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \mathcal{F}(\mathbf{r}_{0,t-1}) & \mathcal{F}(\mathbf{r}_{1,t-1}) & \cdots & I_{n_x\times n_x} \end{bmatrix}$$

$$\mathbf{G}(\mathbf{r}_{0,t}) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C_{r_1} & 0 & \cdots & 0 \\ C_{r_2}\mathcal{F}(\mathbf{r}_1) & C_{r_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ C_{r_t}\mathcal{F}(\mathbf{r}_{1,t-1}) & C_{r_t}\mathcal{F}(\mathbf{r}_{2,t-1}) & \cdots & C_{r_t} \end{bmatrix}.$$

The problem we addressed is as follow: Given a sequence of measurements $\mathbf{y}_{0,t}$, find a recursive state estimation algorithm for the MJLSs that yields optimal estimates of $\mathbf{x}_{0,t}$.

A. Estimation Strategy

In this paper, a new strategy is used to obtain the estimates of $\mathbf{r}_{t-N,t}$ and $\mathbf{x}_{t-N,t}$ at each instant t with $t \geq N > 0$. The ML algorithm is a suitable method for mode detection, since it yields the optimal estimates of mode jumping sequence $\mathbf{r}_{t-N,t}$ in a *maximum a posteriori* sense. On the other hand, based on the optimal estimates of $\mathbf{r}_{t-N,t}$, the MHE is well suited for obtaining the state estimates of $\mathbf{x}_{t-N,t}$ by minimizing a fixed-interval quadratic cost function. The estimation strategy is given in two steps as follows.

1) Maximum Likelihood-Based Mode Detection: The maximum likelihood-based (ML-based) mode detection algorithm is to determine the most likely jumping mode sequence within the interval $[t-N,t]$ with the mode probability function $f(\mathbf{r}_{t-N,t}|x_{t-N}, \mathbf{y}_{t-N,t})$. Let $f(\mathbf{y}_{t-N,t}|x_{t-N}, \mathbf{r}_{t-N,t})$ represent the probability density function of measurement sequence $\mathbf{y}_{t-N,t}$ in the interval $[t-N,t]$ conditioned on x_{t-N} and $\mathbf{r}_{t-N,t}$. Let $\hat{\mathbf{r}}_{t-N,t}$ denote the ML estimate of $\mathbf{r}_{t-N,t}$. Then, $\hat{\mathbf{r}}_{t-N,t}$ can be found by maximizing the log-likelihood function $\ell(\mathbf{r}_{t-N,t}, x_{t-N}|\mathbf{y}_{t-N,t})$ of $\mathbf{r}_{t-N,t}$, expressed as

$$\hat{\mathbf{r}}_{t-N,t} = \arg\max_{\mathbf{r}_{t-N,t}} \left\{ \arg\max_{x_{t-N}} \ell(x_{t-N}, \mathbf{r}_{t-N,t}|\mathbf{y}_{t-N,t}) \right\}. \quad (2)$$

The term $\ell(x_{t-N}, \mathbf{r}_{t-N,t}|\mathbf{y}_{t-N,t})$ is the logarithm of the probability function $f(x_{t-N}, \mathbf{r}_{t-N,t}|\mathbf{y}_{t-N,t})$, and obtainable from

$$\ell(x_{t-N}, \mathbf{r}_{t-N,t}|\mathbf{y}_{t-N,t}) \triangleq \log\{f(x_{t-N}, \mathbf{r}_{t-N,t}|\mathbf{y}_{t-N,t})\}$$

and

$$f(x_{t-N}, \mathbf{r}_{t-N,t}|\mathbf{y}_{t-N,t}) = \frac{1}{a}\Pr(\mathbf{r}_{t-N,t}|\mathfrak{F}) \times f(\mathbf{y}_{t-N,t}, x_{t-N}|\mathbf{r}_{t-N,t})$$

where a is the normalizing constant, and obtainable from

$$a = \sum_{\mathbf{r}_{t-N,t}\in\mathbb{M}^{N+1}} \Pr(\mathbf{r}_{t-N,t}|\mathfrak{F}) \int f(\mathbf{y}_{t-N,t}, x_{t-N}|\mathbf{r}_{t-N,t}) dx_{t-N}$$

with $\Pr(\mathbf{r}_{t-N,t}|\mathfrak{F}) = \Pr(r_{t-N}|\mathfrak{F}) \prod_{s=t-N}^{t} \Pr(r_{s+1}|r_s)$, and the probability distribution of r_{t-N} being a discrete uniform distribution.

2) Moving Horizon Sequence Estimate of $\mathbf{x}_{t-N,t}$: Given a predictor $\bar{x}_{t-N}$ of initial state x_{t-N}, we define the error in the initial predictor as $\bar{e}_{t-N} \triangleq x_{t-N} - \bar{x}_{t-N}$. The moving-horizon state estimation problem for MJLS can be formulated as minimizing a quadratic cost function of errors $\bar{e}_{t-N}$, $\mathbf{w}_{t-N,t-1}$, and $\mathbf{v}_{t-N,t}$. Accordingly, using the optimal estimate $\hat{\mathbf{r}}_{t-N,t}$ obtained from the ML algorithm, the weighted least-squares state estimates of the unknown variables x_{t-N}, $\mathbf{w}_{t-N,t-1}$ can be calculated by solving the following optimization problem:

$$\{\hat{x}_{t-N}, \hat{\mathbf{w}}_{t-N,t-1}\} = \arg\min_{x_{t-N}, \mathbf{w}_{t-N,t-1}} \mathcal{J}(\bar{x}_{t-N}, \mathbf{y}_{t-N,t}, \hat{\mathbf{r}}_{t-N,t}) \quad (3)$$

where

$$\mathcal{J}(\bar{x}_{t-N}, \mathbf{y}_{t-N,t}, \hat{\mathbf{r}}_{t-N,t}) := \|\bar{e}_{t-N}\|_P^2 + \|\mathbf{w}_{t-N,t-1}\|_{\mathcal{Q}}^2 + \|\mathbf{y}_{t-N,t} - \mathbf{H}(\hat{\mathbf{r}}_{t-N,t})x_{t-N} - \mathbf{G}(\hat{\mathbf{r}}_{t-N,t})\mathbf{w}_{t-N,t-1}\|_{\mathcal{R}}^2$$

$$\mathcal{Q} = \mathrm{diag}\underbrace{\{Q_w, Q_w, \dots, Q_w\}}_{N\,\mathrm{blocks}}$$

$$\mathcal{R} = \mathrm{diag}\underbrace{\{R_v, R_v, \dots, R_v\}}_{(N+1)\mathrm{blocks}}$$

and the symmetric matrices $P \in \mathbb{R}^{n_x\times n_x}$, $Q \in \mathbb{R}^{n_w\times n_w}$, and $R \in \mathbb{R}^{n_v\times n_v}$ are designed parameters that satisfy $P > 0$, $Q > 0$, and $R > 0$. Note that P functions to penalize the distance between the true state value x_{t-N} and its prediction $\bar{x}_{t-N}$ at the beginning of the sliding window; Q and R serve as quantitative measures of our confidence in the dynamical system model and the observation model.

It we use $\hat{x}_{t-N}$ and $\hat{\mathbf{w}}_{t-N,t-1}$ to denote the least-squares solutions of (3) for x_{t-N} and $\mathbf{w}_{t-N,t-1}$ in the interval $[t-N,t]$, respectively, then, the following model:

$$\hat{x}_{k+1} = A_{\hat{r}_k}\hat{x}_k + \hat{w}_k, \ \hat{r}_k \in \hat{\mathbf{r}}_{t-N,t}, \ k \in [t-N, t-1] \quad (4)$$

which is based on the system model in (1), can be used to obtain the moving horizon estimates $\hat{\mathbf{x}}_{t-N+1,t}$.

The next section discusses the observability for system Σ.

III. Observability Analysis

In this section, we use the given noise measurements $\mathbf{y}_{0,N}$ within the interval $[0,N]$ to reconstruct the hybrid discrete states $(x_0, \mathbf{r}_{0,N})$, and give the observability conditions under the assumption that no *a priori* probabilistic information is provided.

A. Observability in Noise-Free Case

Consider the noise-free Σ described by

$$\hat{\Sigma}: \begin{cases} \mathbf{x}_{\hat{\Sigma}}(N+1, x_0, \mathbf{r}_{0,N}) = \mathbf{F}(\mathbf{r}_{0,N})\, x_0 \\ \mathbf{y}_{\hat{\Sigma}}(N, x_0, \mathbf{r}_{0,N}) = \mathbf{H}(\mathbf{r}_{0,N})\, x_0. \end{cases} \tag{5}$$

The definitions of the observability for the MJLS in the noise-free case are given below:

Definition 3.1: (Observability) An MJLS $\hat{\Sigma}$ is said to be observable, if there exists a finite positive integer N such that by knowing the measurements $\mathbf{y}_{\hat{\Sigma}}(\mathrm{N}, x_0, \mathbf{r}_{0,N})$ it is sufficient to determine the initial state x_0 and the mode jumping sequence $\mathbf{r}_{0,\mathrm{N}}$.

As shown in [30], an essential result for mode observability is given on high level of generality. Accordingly, a similar definition for mode-distinguishability of the MJLS is given as follows.

Definition 3.2: (Mode-distinguishability) [30] For MJLSs in the noise-free case, the states $(x_0, \mathbf{r}_{0,N})$ and $(\bar{x}_0, \bar{\mathbf{r}}_{0,N})$ are distinguishable in $(N+1)$ steps if $\mathbf{H}(\mathbf{r}_{0,N})x_0 \neq \mathbf{H}(\bar{\mathbf{r}}_{0,N})\bar{x}_0$ holds for any $\mathbf{r}_{0,N} \neq \bar{\mathbf{r}}_{0,N} \in \mathbb{M}^{N+1}$ with $x_0,\ \bar{x}_0 \in \mathbb{R}^{n_x}$ and $x_0 \neq \mathbf{0}$ or $\bar{x}_0 \neq \mathbf{0}$, $\mathbf{0} \in \mathbb{R}^{n_x}$.

If the mode sequence $\mathbf{r}_{0,N} \in \mathbb{M}^{N+1}$ is observable, then system $\hat{\Sigma}$ is observable since there exist a positive integer t, $N \geq \mathrm{t}$, and a scalar γ such that for any $\mathbf{r}_{0,N}$ and any two initial states $x_0, \bar{x}_0 \in \mathbb{R}^{n_x}$ with $x_0 \neq \bar{x}_0$, the following inequality:

$$\gamma\|x_0 - \bar{x}_0\|^2 \leq \|\mathbf{y}_{\hat{\Sigma}}(N, x_0, \mathbf{r}_{0,N}) - \mathbf{y}_{\hat{\Sigma}}(N, \bar{x}_0, \mathbf{r}_{0,N})\|^2 \tag{6}$$

is equivalent to $\|\mathbf{H}(\mathbf{r}_{0,N})x_0 - \mathbf{H}(\mathbf{r}_{0,N})\bar{x}_0\| \geq \gamma\|x_0 - \bar{x}_0\|$. Therefore, the observability matrix can be defined as $\mathbf{H}(\mathbf{r}_{0,N})$ for $t \in \mathbb{T}_0^N$, and the inequality in (6) can be satisfied when $\mathbf{H}(\mathbf{r}_{0,N})$ is full rank for all $\mathbf{r}_{0,N} \in \mathbb{M}^{N+1}$.

From Definition 3.2, a sufficient condition for mode distinguishability of the systems $(x_0, \mathbf{r}_{0,N})$ and $(\bar{x}_0, \bar{\mathbf{r}}_{0,N})$ in the interval $[0, N]$ is

$$\mathbf{H}(\mathbf{r}_{0,N})x_0 \neq \mathbf{H}(\bar{r}_{0,N})\bar{x}_0$$

which can be rewritten as

$$[\mathbf{H}(\mathbf{r}_{0,N})\ \mathbf{H}(\bar{\mathbf{r}}_{0,N})] \begin{bmatrix} x_0 \\ -\bar{x}_0 \end{bmatrix} \neq \mathbf{0}. \tag{7}$$

For (7) to hold, it is evident that sufficient condition of distinguishability reduces to $\operatorname{rank}([\mathbf{H}(\mathbf{r}_{0,N}),\ \mathbf{H}(\bar{\mathbf{r}}_{0,N})]) = 2n_x$. Therefore, by define the *joint observability matrix* as $\mathcal{O}(\mathbf{r}_{0,N}, \bar{\mathbf{r}}_{0,N}) = [\mathbf{H}(\mathbf{r}_{0,N}),\ \mathbf{H}(\bar{\mathbf{r}}_{0,N})]$, system $\hat{\Sigma}$ is mode distinguishable and *joint observable* if there exist a positive integer $\bar{\mathrm{t}}_0$ and $N \geq \bar{\mathrm{t}}_0$, such that $\operatorname{rank}(\mathcal{O}(\mathbf{r}_{0,N}, \bar{\mathbf{r}}_{0,N})) = 2n_x$ for all $\mathbf{r}_{0,N} \neq \bar{\mathbf{r}}_{0,N} \in \mathbb{M}^{N+1}$.

However, it has been shown in [20] that if any two mode jumping sequences $\mathbf{r}_{0,N} \neq \bar{\mathbf{r}}_{0,N}$, $N \geq \bar{\mathrm{t}}_0$, differ only in the first or in the last $\lceil n_x/n_y \rceil - 1$ instants of the observations window $[0, N]$, then the joint observability condition $\operatorname{rank}(\mathcal{O}(\mathbf{r}_{0,N}, \bar{\mathbf{r}}_{0,N})) = 2n_x$ cannot be satisfied. On the basis of this result, if given $n_x > n_y$, the mode observability is confined to a restricted interval $[\alpha, N-\beta]$, in which the positive integers α, β satisfy $0 < \alpha \leq N - \beta$, then, the following lemma is applicable for the case when the system is (α, β)-mode observable.

Lemma 3.1: [20] An MJLS $\hat{\Sigma}$ is (α, β)-mode observable in the $N+1$ steps for any pair of jumping sequences $\mathbf{r}_{0,N}, \bar{\mathbf{r}}_{0,N} \in \mathbb{M}^{N+1}$ with $\mathbf{r}_{\alpha,N-\beta} \neq \bar{\mathbf{r}}_{\alpha,N-\beta} \in \mathbb{R}^{N-\alpha-\beta+1}$, if their joint observability matrix satisfies the condition $\operatorname{rank}(\mathcal{O}(\mathbf{r}_{0,N}, \bar{\mathbf{r}}_{0,N})) = 2n_x$.

The study for the noise-free case presented above is taken as a starting point to analyze the stochastic observability for system Σ.

B. Observability in the Presence of Gaussian Noise

Stochastic observability is widely used to design asymptotically stable filters for the discrete-time MJLS subjected to independent Gaussian perturbations.

Consider system Σ. Under the second condition given in Assumption 2.1, we set

$$\bar{\mathbf{y}}_{\Sigma}(N, x_0, \mathbf{r}_{0,N}) = \mathrm{E}\{\mathbf{y}|x_0, \mathbf{r}_{0,N}\} = \mathbf{H}(\mathbf{r}_{0,N})x_0. \tag{8}$$

Consequently, the following result of the stochastic observability for system Σ is proposed purely as a technical condition in the proof of the stability of the MHE-scheme algorithm.

Definition 3.3: (Incrementally stochastic observable) An MJLS Σ is incrementally stochastic observable, if there exist integer $\bar{\mathrm{t}}$ and constant $\bar{\gamma}$ such that for arbitrary Markov chain $(\mathbf{r}_{0,N}, \Pi)$ and any $x_0, \bar{x}_0 \in \mathbb{R}^{n_x}$ with $x_0 \neq \bar{x}_0$, the following inequality:

$$\bar{\gamma}\|x_0 - \bar{x}_0\|^2 \leq \|\bar{\mathbf{y}}_{\Sigma}(N, x_0, \mathbf{r}_{0,N}) - \bar{\mathbf{y}}_{\Sigma}(N, \bar{x}_0, \mathbf{r}_{0,N})\|^2 \tag{9}$$

holds for all $N \geq \bar{\mathrm{t}}$.

From Definition 3.3, the right-hand side of (9) can be rewritten as $\|\mathbf{H}(\mathbf{r}_{0,N})(x_0 - \bar{x}_0)\|^2$.

Therefore, the observability conditions and properties given in Definition 3.2 and Lemma 3.1 are also applicable to analyze the stochastic observability for the MJLS in the presence of Gaussian noise.

Definition 3.4: The MJLSs in the presence of Gaussian noise with arbitrary Markov chain $(\mathbf{r}_{0,N}, \Pi)$ is (α, β)-mode stochastic observable in the $N+1$ steps for any pair of $\mathbf{r}_{0,N}, \bar{\mathbf{r}}_{0,N} \in \mathbb{M}^{N+1}$ with $\mathbf{r}_{\alpha,N-\beta} \neq \bar{\mathbf{r}}_{\alpha,N-\beta} \in \mathbb{M}^{N-\alpha-\beta+1}$, if their joint observability matrix $\operatorname{rank}(\mathcal{O}(\mathbf{r}_{0,N}, \bar{\mathbf{r}}_{0,N})) = 2n_x$.

On the basis of a sufficient stochastic observable condition provided by Definition 3.4, a mode estimation technique is proposed for the system Σ based on the probability measure technique. The reference probability method is used to construct the information states for taking into account specific mode sequence $\mathbf{r} \in \mathbb{M}^{N+1}$, $N > \mathrm{t}_0$, and the observing sequence $\mathbf{y}_{0,N}$ in the interval $[0, N]$. Accordingly, [27] presented an innovative mode estimation algorithm for switching systems subject to Gaussian disturbance. The basic principle is to estimate the system modes based on the maximum likelihood procedure, which is called "Bayesian statistical correction," see [31] for details. In the next section, we use this mode estimation technique to construct a mode detection algorithm for system Σ.

C. Mode Detection Algorithm

For any $\bar{\mathbf{r}}, \mathbf{r} \in \mathbb{M}^{N+1}$, $N > \mathrm{t}_0$, the probability measure $\mathbf{Pr}(\bar{\mathbf{r}} \succ \mathbf{r}|x_0, \tilde{\mathbf{r}})$ expresses that the mode sequence $\bar{\mathbf{r}}$ is closer than $\mathbf{r}$ to the true mode sequence $\tilde{\mathbf{r}}$, given the true initial state x_0. Since w_k and v_k, $k \in [0, N]$ obey the Gaussian distribution, the probability density function $f(\mathbf{y}_{0,N}, x_0|\mathbf{r})$ can be computed as a Gaussian distribution as follows:

$$f(\mathbf{y}_{0,N}, x_0|\mathbf{r}) = \phi\left(\mathbf{y}_{0,N} - \mathbf{H}(\mathbf{r})x_0; \mathbf{0}_{(N+1)n_y}, \Delta(\mathbf{r})\right).$$

According to Assumption 2.1, covariance $\Delta(\mathbf{r})$ is defined as

$$\begin{aligned}\Delta(\mathbf{r}) &= \mathrm{E}\{\mathbf{e}(\mathbf{r})\mathbf{e}(\mathbf{r})'\}\\ &= E\{(\mathbf{y}_{0,N} - \mathbf{H}(\mathbf{r})x_0)(\cdot)'\}\\ &= \mathrm{E}\{\mathbf{G}(\mathbf{r})\mathbf{w}_{0,N-1}\mathbf{w}_{0,N-1}{}'\mathbf{G}(\mathbf{r})' + \mathbf{v}_{0,N}\mathbf{v}_{0,N}'\}\\ &= \mathbf{G}(\mathbf{r})\mathrm{diag}\underbrace{\{Q_w, Q_w, \ldots, Q_w\}}_{N\,\mathrm{block}}\mathbf{G}(\mathbf{r})'\\ &\quad + \mathrm{diag}\underbrace{\{R_v, R_v, \ldots, R_v\}}_{(N+1)\,\mathrm{block}}.\end{aligned}$$

To establish a connection between $\mathbf{Pr}(\bar{\mathbf{r}} \succ \mathbf{r}|x_0, \tilde{\mathbf{r}})$ and the mode-conditional probability density function $f(\mathbf{y}_{0,N}, x_0|\mathbf{r})$, a new probability measure $\mathbf{Pr}(\bar{\mathbf{r}} \succ \mathbf{r}|x_0, \tilde{\mathbf{r}})$ is constructed based on the following inequality:

$$f(x_0, \bar{\mathbf{r}}|\mathbf{y}_{0,N}) > f(x_0, \mathbf{r}|\mathbf{y}_{0,N})$$

where the likelihood function $f(x_0, \mathbf{r}|\mathbf{y}_{0,N}) \propto \Pr(\mathbf{r}|\mathfrak{F})f(\mathbf{y}_{0,N}, x_0|\mathbf{r})$. The predicted probability of mode sequence $\mathbf{r} = \{r_0, r_1, \ldots, r_N\}$ can be calculated from

$$\Pr(\mathbf{r}|\mathfrak{F}) = \prod_{k=1}^{N} \Pr(r_k|r_{k-1}, \mathfrak{F})\Pr(r_0|\mathfrak{F}).$$

When x_0 and $\mathbf{r}$ are available, the predicted measurements vector is

$$\bar{\mathbf{y}}_{0,N}(\mathbf{r}) = \mathbf{H}(\mathbf{r})x_0.$$

Define the residual error as $\mathbf{e}(\mathbf{r}) = \mathbf{y}_{0,N} - \bar{\mathbf{y}}_{0,N}(\mathbf{r}) \sim \Phi(0, \Delta(\mathbf{r}))$ with respect to the mode sequence $\mathbf{r} \in \mathbb{M}^{N+1}$. Given the true initial state x_0, we define its corresponding observation merit function as

$$f(\mathbf{y}_{0,N}, x_0|\mathbf{r}) = \frac{1}{\sqrt{|2\boldsymbol{\pi}\Delta(\mathbf{r})|}}\exp\left(-\frac{1}{2}e(\mathbf{r})\left(\Delta(\mathbf{r})\right)^{-1}e(\mathbf{r})'\right).$$

Then, based on Bayesian theorem, the likelihood probability function for updating the mode probabilities is given by

$$f(x_0, \mathbf{r}|\mathbf{y}_{0,N}) = \frac{1}{a_1}f(\mathbf{y}_{0,N}, x_0|\mathbf{r})\Pr(\mathbf{r}|\mathfrak{F}) \tag{10}$$

where a_1 is a normalized parameter. Hence, given the initial state x_0, the optimal estimate $\hat{\mathbf{r}}$

$$\begin{aligned}\hat{\mathbf{r}} &\propto \arg\max_{\mathbf{r}\in\mathbb{M}^{N+1}}\left(\max_{x_0\in\mathbb{R}^{n_x}} f(x_0, \mathbf{r}|\mathbf{y}_{0,N})\right)\\ &\propto \arg\max_{\mathbf{r}\in\mathbb{M}^{N+1}}\left(\arg\max_{x_0\in\mathbb{R}^{n_x}}(\ell(x_0, \mathbf{r}|\mathbf{y}_{0,N}))\right)\\ &= \arg\max_{\mathbf{r}\in\mathbb{M}^{N+1}}\left(\arg\max_{x_0\in\mathbb{R}^{n_x}}(\log f(\mathbf{y}_{0,N}, x_0|\mathbf{r}))\right.\\ &\qquad\left. + \log\Pr(\mathbf{r}|\mathfrak{F})\right).\end{aligned} \tag{11}$$

For the term $f(\mathbf{y}_{0,N}, x_0|\mathbf{r})$, we have

$$\begin{aligned}&\max_{x_0\in\mathbb{R}^{n_x}} f(\mathbf{y}_{0,N}, x_0|\mathbf{r})\\ &= \max_{x_0\in\mathbb{R}^{n_x}} \phi\left(\mathbf{y}_{0,N} - \mathbf{H}(\mathbf{r})x_0; \mathbf{0}_{(N+1)n_y}, \Delta(\mathbf{r})\right)\\ &= \max_{x_0\in\mathbb{R}^{n_x}} \frac{1}{\sqrt{|2\boldsymbol{\pi}\Delta(\mathbf{r})|}}\exp\left(-\frac{1}{2}d(\mathbf{y}_{0,N}, x_0, \mathbf{r})\right)\end{aligned} \tag{12}$$

where $d(\mathbf{y}_{0,N}, \mathbf{r}, x_0)$ is defined as

$$d(\mathbf{y}_{0,N}, \mathbf{r}, x_0) \triangleq \|\mathbf{y}_{0,N} - \mathbf{H}(\mathbf{r})x_0\|^2_{\Delta(\mathbf{r})^{-1}}.$$

Denote the minimum distance by $\bar{d}(\mathbf{y}_{0,N}, \mathbf{r})$, which is given as

$$\bar{d}(\mathbf{y}_{0,N}, \mathbf{r}) \triangleq \min_{x_0\in\mathbb{X}} d(\mathbf{y}_{0,N}, \mathbf{r}, x_0).$$

Then, a generalized maximum-likelihood criterion for the mode stochastic observability of the MJLS is presented as follows.

Theorem 3.1: Consider an MJLS Σ. If for any pair of mode jumping sequences $\mathbf{r} \neq \bar{\mathbf{r}} \in \mathbb{M}^{N+1}$, $\mathrm{rank}\left(\mathcal{O}(\bar{\mathbf{r}}, \mathbf{r})\right) = 2n_x$ holds, then system Σ is (α, β)-mode stochastic observable, and estimate $\hat{\mathbf{r}}_{0,N}$ can be obtained by solving the minimization problem

$$\begin{aligned}\hat{\mathbf{r}}_{0,N} \in \arg\min_{\mathbf{r}\in\mathbb{M}^{N+1}} &\left\{-2\log\{\Pr(\mathbf{r})\} + \log\{|\Delta(\mathbf{r})|\}\right.\\ &\left.+ \bar{d}(\mathbf{y}_{0,N}, \mathbf{r})\right\}\end{aligned} \tag{13}$$

where $N \geq \alpha + \beta + 1$.

The proof of Theorem 3.1 is shown in Appendix A.

From Theorem 3.1, we observe that $\hat{\mathbf{r}}_{0,N}$ is a more reliable estimate of the mode jumping sequence within the time interval $[0, N]$. The following theorem confirms that the estimate $\hat{\mathbf{r}}_{0,t}$ obtained from Theorem 3.1 is optimal in the maximum-likelihood sense.

Theorem 3.2: Assume the true mode jumping sequence within the time interval $[0, N]$ is given by $\tilde{\mathbf{r}}$. If, for any $\mathbf{r} \in \mathbb{M}^{N+1} \neq \tilde{\mathbf{r}}$, $\mathrm{rank}\left(\mathcal{O}(\tilde{\mathbf{r}}, \mathbf{r})\right) = 2n_x$ holds, there exist two suitable scalars $\kappa(\tilde{\mathbf{r}}, \mathbf{r}) > 0$ and $\varsigma(\tilde{\mathbf{r}}, \mathbf{r}) > 0$ such that the following inequality holds:

$$\begin{aligned}\mathbf{Pr}(\tilde{\mathbf{r}} \succ \mathbf{r}|x_0, \tilde{\mathbf{r}}) \geq \gamma_{(N+1)n_y}&\left(\mathbf{z}'\Delta(\tilde{\mathbf{r}})\mathbf{z} \leq \frac{2}{\varsigma(\tilde{\mathbf{r}}, \mathbf{r})}\log\frac{\Pr(\tilde{\mathbf{r}})}{\Pr(\mathbf{r})}\right.\\ &\left.+ \frac{1}{\varsigma(\tilde{\mathbf{r}}, \mathbf{r})}\log\frac{|\Delta(\mathbf{r})|}{|\Delta(\tilde{\mathbf{r}})|} + \frac{\kappa(\tilde{\mathbf{r}}, \mathbf{r})}{\varsigma(\tilde{\mathbf{r}}, \mathbf{r})}\|x_0\|^2\right)\end{aligned}$$

then

$$\lim_{\|x_0\|^2\to\infty}\mathbf{Pr}(\tilde{\mathbf{r}} \succ \mathbf{r}|x_0, \tilde{\mathbf{r}}) = 1.$$

The proof of Theorem 3.2 is given in Appendix B. ■

In the next section, the mode observability results are applied to develop a moving horizon state estimator scheme for the MJLSs in the presence of Gaussian noise.

IV. Design of Moving Horizon Estimator

In Section II, a moving horizon estimator scheme is given for yielding moving horizon state estimates as the "best fit" of the state trajectory of the system Σ dynamic process. In Section III, a mode detection method is proposed for obtaining the optimal mode estimates for system Σ in the maximum *a posteriori* sense. We now proceed to derive the mode detection-moving horizon estimation (MD-MHE) algorithm.

A. Mode Detection-Moving Horizon Estimation

According to Theorem 3.1, given measurements $\mathbf{y}_{t-N,t}$, a maximum likelihood criterion to obtain the estimate of the mode sequence state within the horizon $[t-N,t]$ is given as

$$\hat{\mathbf{r}}_{t-N,t} = \arg\min_{\mathbf{r}\in\mathbb{M}^{N+1}} \{-2\log\{P(\mathbf{r}|\mathfrak{F})\} + \log|\Delta(\mathbf{r})| + d(\mathbf{y}_{t-N,t},\mathbf{r})\} \tag{14}$$

where the estimates of $\mathbf{r}_{t-N,t}$ obtained at t are denoted by $\hat{\mathbf{r}}_{t-N,t}$. Recall Definition 3.4, we find that the term $\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t} \triangleq \{\hat{r}_{t-N+\alpha|t}, \hat{r}_{t-N+\alpha+1|t}, \ldots, \hat{r}_{t-\beta|t}\} \subseteq \hat{\mathbf{r}}_{t-N,t}$ is more reliable in the time interval $[t-N+\alpha, t-\beta]$.

First, define $\hat{x}_{t-N-1+\alpha}$ and $\hat{\mathbf{w}}^{\alpha,\beta}_{t-N-1,t-1} = \{\hat{w}_{t-N-1+\alpha|t-1}, \hat{w}_{t-N+\alpha|t-1}, \ldots, \hat{w}_{t-N-1-\beta|t-1}\}$ as the MHE-based estimates of $x_{t-N-1+\alpha}$ and $\mathbf{w}^{\alpha,\beta}_{t-N-1,t-1}$ at $t-1$. Next, we derive the arrival cost, which summarizes the effect of the information on the states $\mathbf{x}^{\alpha,\beta}_{t-N,t}$. Since the prediction $\bar{x}_{t-N+\alpha}$ is the crux of the formulation of arrival cost, it can be computed based on the estimates $\hat{x}_{t-N-1+\alpha|t-1}$, and $\hat{r}_{t-N-1+\alpha|t-1}$ via a system dynamic function, which is given as

$$\bar{x}_{t-N+\alpha} = A_{\hat{r}_{t-N-1+\alpha|t-1}} \hat{x}_{t-N-1+\alpha|t-1} + \hat{w}_{t-N-1+\alpha|t-1}. \tag{15}$$

After obtaining the optimal mode sequence estimate $\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t}$ by (14) and the prediction $\bar{x}_{t-N+\alpha}$ by (15), the objective is to find the estimates of the continuous state vectors $\mathbf{x}^{\alpha,\beta}_{t-N,t}$ from measurements $\mathbf{y}_{t-N,t}$ in the least-squares sense. Hence, an objective function is introduced as

$$\begin{aligned} &J\left(\bar{x}_{t-N+\alpha}, \hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t}, \mathbf{y}^{\alpha,\beta}_{t-N,t}\right) \\ &= \|x_{t-N+\alpha} - \bar{x}_{t-N+\alpha}\|^2_P + \|\mathbf{w}^{\alpha,\beta}_{t-N,t}\|^2_{\tilde{\mathcal{Q}}} \\ &+ \|\mathbf{y}^{\alpha,\beta}_{t-N,t} - \mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\bar{x}_{t-N+\alpha} - \mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\mathbf{w}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{R}}} \end{aligned}$$

where

$$\tilde{\mathcal{Q}} = \operatorname{diag}\left\{Q^{(1)}_w, Q^{(2)}_w, \ldots, Q^{(N-\alpha-\beta)}_w\right\}$$

$$\tilde{\mathcal{R}} = \operatorname{diag}\left\{R^{(1)}_v, R^{(2)}_v, \ldots, R^{(N-\alpha-\beta+1)}_v\right\}.$$

with $Q^{(i)}_w$ and $R^{(i)}_v$ being the ith components of $\tilde{\mathcal{Q}}$ and $\tilde{\mathcal{R}}$, respectively.

We now present the estimation algorithm.

Mode detection-moving horizon estimation.

Initialization: Given measurements $\mathbf{y}_{0,N}$, find the ML-estimates of $\mathbf{r}_{0,N}$ by solving

$$\hat{\mathbf{r}}_{0,N} = \arg\min_{\mathbf{r}_{0,N}\in\mathbb{M}^{N+1}} \{-2\log P(\mathbf{r}_{0,N}|\mathfrak{F}) + \log|\Delta(\mathbf{r}_{0,N})| + d(\mathbf{y}_{0,N},\mathbf{r})\}.$$

Then, $\bar{x}_\alpha \approx \mathcal{F}(\hat{\mathbf{r}}_{0,\alpha-1})\bar{x}_0$.

Step 1: The MD-MHE algorithm is stated at any time $t \in \{N, N+1, \ldots\}$.

1) Given measurements $\mathbf{y}_{t-N,t}$, find the estimates of $\mathbf{r}_{t-N,t}$ by solving

$$\hat{\mathbf{r}}_{t-N,t} = \arg\min_{\mathbf{r}_{t-N,t}} \{-2\log P(\mathbf{r}_{t-N,t}|\mathfrak{F}) + \log|\Delta(\mathbf{r}_{t-N,t})| + d(\mathbf{y}_{t-N,t},\mathbf{r})\}.$$

2) Given the prediction $\bar{x}_{t-N+\alpha}$, find the solution of the minimization problem

$$\left\{\hat{x}_{t-N+\alpha|t}, \hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\right\} \triangleq \arg\min_{x_{t-N+\alpha}, \mathbf{w}^{\alpha,\beta+1}_{t-N,t}} J\left(\bar{x}_{t-N+\alpha}, \hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t}, \mathbf{y}^{\alpha,\beta}_{t-N,t}\right).$$

3) The optimal state estimates of $\mathbf{x}^{\alpha+1,\beta}_{t-N,t}$ at t are obtained from

$$\hat{\mathbf{x}}^{\alpha+1,\beta}_{t-N,t} = \mathbf{F}\left(\hat{\mathbf{r}}^{\alpha,\beta+1}_{t-N,t}\right)\hat{x}_{t-N+\alpha|t} + \mathbf{B}\left(\hat{\mathbf{r}}^{\alpha,\beta+1}_{t-N,t}\right)\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}.$$

Step 2: The prediction for the next cycle of moving horizon estimator is propagated as

$$\bar{x}_{t-N+1+\alpha|t} = A_{\hat{r}_{t-N+\alpha|t}} \hat{x}_{t-N+\alpha|t} + \hat{w}_{t-N+\alpha|t}$$
$$\hat{r}_{t-N+\alpha|t} \in \hat{\mathbf{r}}_{t-N,t}.$$

Step 3: Let $t = t+1$, and go to *Step 1*.

The block diagram of the MD-MHE algorithm scheme is shown in Fig. 1. Note that in the MD-MHE algorithm, the MH strategy is operated in the window $[t-N+\alpha, t-\beta]$ to provide the estimates $\hat{\mathbf{x}}^{\alpha,\beta}_{t-N,t}$. However, there exists a certain delay β when obtaining an optimal estimate of the state x_t, while a part of the mode detection results obtained in the window $[t-N,t]$ is under-utilized in the MD-MHE algorithm. We observe that the mode estimates $\hat{\mathbf{r}}^{N-\beta+1,0}_{t-N,t}$ is completely ignored. Thus, by using $\hat{\mathbf{r}}^{N-\beta+1,0}_{t-N,t}$, the exploitation of the moving horizon state estimates is extended to the restricted interval $[t-\beta+1, t]$, and the time delay in the MD-MHE algorithm can be eliminated.

Using Lemma 3.1, we divide the estimates of the current mode jumping sequence into three parts as

$$\hat{\mathbf{r}}_{t-N,t} = \left\{\hat{\mathbf{r}}^{0,N-\alpha+1}_{t-N,t} \oplus \hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t} \oplus \hat{\mathbf{r}}^{N-\beta+1}_{t-N,t}\right\}. \tag{16}$$

The mode sequence estimates $\mathbf{r}^{\alpha,\beta}_{t-N+,t}$ are fully reliable because the system satisfies the conditions of the (α,β)-mode observability. Although the mode estimate $\hat{\mathbf{r}}^{N-\beta+1,0}_{t-\beta+1,t}$ provided by the maximum likelihood algorithm is not completely reliable according to Definition 3.4, its reliability is still higher than other candidates $\mathbf{r}^{N-\beta+1,0}_{t-N,t} \neq \hat{\mathbf{r}}^{N-\beta+1,0}_{t-N,t}$, $\mathbf{r}^{N-\beta+1,0}_{t-N,t} \in \mathbb{M}^\beta$.

For expressing the level of reliability of the estimates, a factor $0 < \zeta \le 1$ is defined to express the degree of reliability of the estimates $\hat{\mathbf{r}}^{N-\beta+1,0}_{t-N,t}$ with $\zeta_{\max} = 1$ for full reliability. Hence, a natural criterion can be derived to measure the influence of the reliability of the estimates $\hat{\mathbf{r}}^{N-\beta+1,0}_{t-N,t}$ on the operation of MH-strategy within the interval $[t-\beta+1, t]$. The optimal terms regarding the states $\mathbf{x}^{N-\beta+1,0}_{t-N,t}$ can be designed using the scalar ζ and adding the states to the objective function. Then, an extended

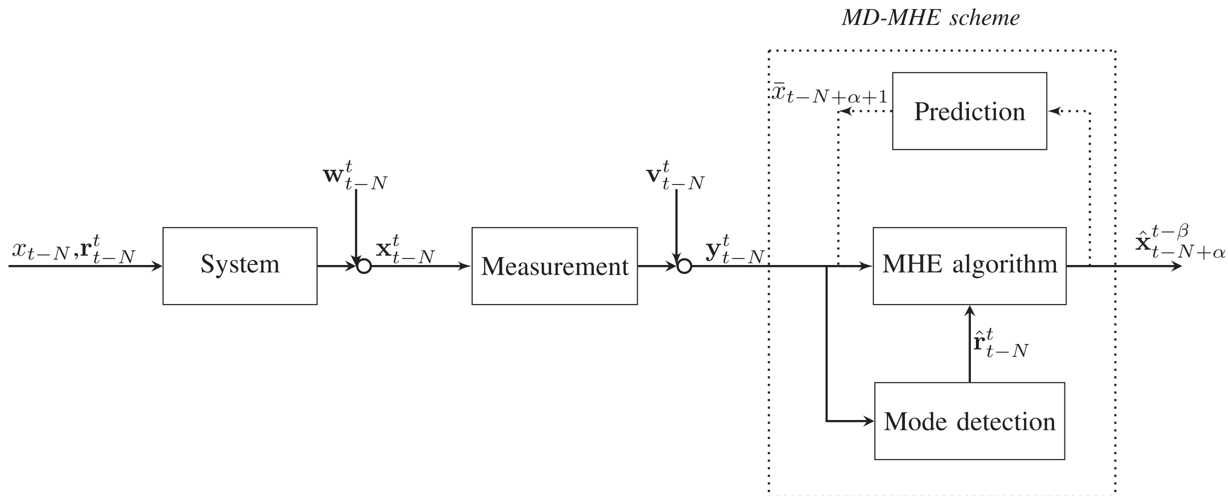


Fig. 1. Mode detection-moving horizon estimation scheme.

objective function on window $[t-N+\alpha, t]$ can be obtained as

$$J\left(\bar{x}_{t-N+\alpha}, \hat{\mathbf{r}}_{t-N,t}^{\alpha,0}, \mathbf{y}_{t-N,t}^{\alpha,0}, \zeta\right) = \|x_{t-N+\alpha} - \bar{x}_{t-N+\alpha}\|_P^2 + \|\mathbf{w}_{t-N,t}^{\alpha,1}\|_{\overline{\mathcal{Q}}}^2 + \|\mathbf{y}_{t-N,t}^{\alpha,0} - \mathbf{H}(\hat{\mathbf{r}}_{t-N,t}^{\alpha,0})\bar{x}_{t-N+\alpha} - \mathbf{G}(\hat{\mathbf{r}}_{t-N,t}^{\alpha,0})\mathbf{w}_{t-N,t}^{\alpha,1}\|_{\overline{\mathcal{R}}}^2 \tag{17}$$

where $\overline{\mathcal{Q}} = \text{diag}[\tilde{\mathcal{Q}}, \hat{\mathcal{Q}}]$, $\overline{\mathcal{R}} = \text{diag}[\tilde{\mathcal{R}}, \hat{\mathcal{R}}]$ with

$$\hat{\mathcal{Q}} = \text{diag}\left\{Q_w^{(1)}, Q_w^{(2)}, \ldots, Q_w^{(\beta)}\right\}$$
$$\hat{\mathcal{R}} = \zeta\text{diag}\left\{R_v^{(1)}, R_v^{(2)}, \ldots, R_v^{(\beta)}\right\}.$$

The scalar ζ is designed to enforce the stability of the estimator, which is given in the stability analysis in the next section.

Based on the analysis above, the construction of the new extended objective function makes it practical for the MHE procedure to be extended to cover the window $[t-N+\alpha, t]$. For the initial prediction, it can be computed via the same mechanism as the MD-MHE estimator. The EMD-MHE is summarized as follows.

Extended mode detection-moving horizon estimation.

Initialization: The initialization step is the same as that of the MD-MHE algorithm.

Step 1: For $t > N$ do:

1) (Identical to MD-MHE) Given measurements $\mathbf{y}_{t-N,t}$, find the estimates of $\mathbf{r}_{t-N,t}$ from

$$\hat{\mathbf{r}}_{t-N,t} = \arg\min_{\mathbf{r}_{t-N,t}} \left\{-2\log P(\mathbf{r}_{t-N,t}|\mathfrak{F}) + \log|\Delta(\mathbf{r}_{t-N,t})| + d(\mathbf{y}_{t-N,t}, \mathbf{r})\right\}.$$

2) Given the prediction $\bar{x}_{t-N+\alpha}$, find the estimates

$$\left\{\hat{x}_{t-N+\alpha|t}, \hat{\mathbf{w}}_{t-N,t}^{\alpha,1}\right\}$$

by solving

$$\left\{\hat{x}_{t-N+\alpha|t}, \hat{\mathbf{w}}_{t-N,t}^{\alpha,1}\right\} \triangleq \arg\min_{x_{t-N+\alpha}, \mathbf{w}_{t-N,t}^{\alpha,1}} J\left(\bar{x}_{t-N+\alpha}, \hat{\mathbf{r}}_{t-N,t}^{\alpha,0}, \mathbf{y}_{t-N,t}^{\alpha,0}, \zeta\right).$$

3) (Identical to MD-MHE) The optimal state estimates of $\mathbf{x}_{t-N,t}^{\alpha+1,0}$ at t can be computed from

$$\hat{\mathbf{x}}_{t-N,t}^{\alpha+1,0} = \mathbf{F}\left(\hat{\mathbf{r}}_{t-N,t}^{\alpha,1}\right)\hat{x}_{t-N+\alpha|t} + \mathbf{B}\left(\hat{\mathbf{r}}_{t-N,t}^{\alpha,1}\right)\hat{\mathbf{w}}_{t-N,t}^{\alpha,1}.$$

Step 2: (Identical to MD-MHE) The prediction for the next cycle of moving horizon estimator is propagated as

$$\bar{x}_{t-N+1+\alpha} = A_{\hat{r}_{t-N+\alpha|t}}\hat{x}_{t-N+\alpha|t} + \hat{w}_{t-N+\alpha|t}$$
$$\hat{r}_{t-N+\alpha|t} \in \hat{\mathbf{r}}_{t-N,t}^{\alpha,0}.$$

Step 3: Let $t = t+1$, and go to *Step 1*.

As a result, the estimates $\hat{\mathbf{r}}_{t-N,t}$ computed by the maximum likelihood method can be fully utilized via the definition of scalar ζ in the EMD-MHE algorithm, and a moving horizon state estimate $\hat{x}_{t|t}$ can be obtained without any delay from measurements $\mathbf{y}_{t-N,t}$.

Note that the computational load of the MHE algorithm grows linearly with the length of the horizon. It is important for the MHE-scheme algorithm to use a smaller horizon length for solving the MHE problem. Since the MD-MHE algorithm may suffer from the time delay, and the EMD-MHE algorithm may suffer from the heavy computational load, another algorithm called the incomplete mode detection MHE-based algorithm is formulated on the horizon $[t-\beta, t]$ using the incomplete mode detection result. The algorithm is capable of eliminating delay and reducing the computational load.

Based on the analysis above, given measurements $\mathbf{y}_{t-\beta,t}$, we can compute the incomplete but feasible estimates $\hat{\mathbf{r}}_{t-N,t}^{N-\beta,0} \in \mathbb{R}^{\beta+1}$ of the jumping sequence via the maximum likelihood algorithm within the restricted interval $[t-\beta, t]$. Presented below is the algorithm.

Incomplete mode detection-moving horizon estimation (IMD-MHE) algorithm

Initialization: Given measurements $\mathbf{y}_{0,N}$, find the ML-estimates of $\mathbf{r}_{0,N}$ by solving

$$\hat{\mathbf{r}}_{0,N} = \arg \min_{\mathbf{r}_{0,N} \in \mathbb{M}^{N+1}} \{-2 \log P(\mathbf{r}_{0,N}|\mathfrak{F}) + \log |\Delta(\mathbf{r}_{0,N})| + d(\mathbf{y}_{0,N}, \mathbf{r})\}.$$

Then we obtain $\bar{x}_{N-\beta} \approx \mathcal{F}(\hat{\mathbf{r}}_{0,N-\beta-1})\bar{x}_0$.

Step 1: For $t > N$ do:

1) (Identical to MD-MHE) Given the measurements $\mathbf{y}_{t-N,t}$, find the estimates of $\mathbf{r}_{t-N,t}$ from

$$\hat{\mathbf{r}}_{t-N,t} = \arg \min_{\mathbf{r}_{t-N,t} \in \mathbb{M}^{N+1}} \{-2 \log P(\mathbf{r}_{t-N,t}|\mathfrak{F}) + \log |\Delta(\mathbf{r}_{t-N,t})| + d(\mathbf{y}_{t-N,t}, \mathbf{r})\}.$$

2) Given the prediction $\bar{x}_{t-\beta}$ and $\hat{\mathbf{r}}_{t-N,t}^{N-\beta,0} \subseteq \hat{\mathbf{r}}_{t-N,t}$, find the estimates $\left\{\hat{x}_{t-\beta|t}, \hat{\mathbf{w}}_{t-N,t}^{N-\beta,0}\right\}$ by solving

$$\arg \min_{x_{t-\beta}, \mathbf{w}_{t-\beta}^{t-1}} \|x_{t-\beta} - \bar{x}_{t-\beta}\|_P^2 + \|\mathbf{w}_{t-N,t}^{N-\beta,1}\|_{\mathcal{Q}}^2 + \|\mathbf{y}_{t-\beta,t} - \mathbf{H}\left(\hat{\mathbf{r}}_{t-N,t}^{N-\beta,0}\right) x_{t-\beta} - \mathbf{G}\left(\hat{\mathbf{r}}_{t-N,t}^{N-\beta,0}\right) \mathbf{w}_{t-N,t}^{N-\beta,0}\|_{\mathcal{R}}^2.$$

3) (Identical to MD-MHE) The optimal state estimates of $\mathbf{x}_{t,t-N}^{N-\beta+1,0}$ at t can be computed from

$$\hat{\mathbf{x}}_{t-N,t}^{N-\beta+1,0} = \mathbf{F}\left(\hat{\mathbf{r}}_{t-N,t}^{N-\beta,1}\right) \hat{x}_{t-\beta|t} + \mathbf{B}\left(\hat{\mathbf{r}}_{t-N,t}^{N-\beta,1}\right) \hat{\mathbf{w}}_{t-N,t}^{N-\beta,1}.$$

Step 2: (Identical to MD-MHE) The prediction for the next cycle of moving horizon estimator is propagated as

$$\bar{x}_{t-\beta+1} = A_{\hat{r}_{t-\beta|t}} \hat{x}_{t-\beta|t} + \hat{w}_{t-\beta|t}, \qquad \hat{r}_{t-\beta|t} \in \hat{\mathbf{r}}_{t-N,t}.$$

Step 3: Let $t = t + 1$, and go to *Step 1*.

Using the IMD-MHE algorithm, we can significantly reduce the computational load by reducing the horizon length of the estimation algorithm. Moreover, the parameter ζ can be adjusted to guarantee the stability of the algorithm.

V. Stability Analysis of Algorithms

In this paper, a mode detection-moving horizon state estimation algorithm has been established by combining the maximum-likelihood approach and the MHE algorithm. For completeness, three different estimators (the MD-MHE, the EMD-MHE, and the incomplete MD-MHE) have been constructed with their own distinguishing features. Therefore, the rigor of their sufficient conditions required for the stability of the state estimators varies. Hence, we specifically analyze their stability in this section. First, we define the following series of quantities for system Σ:

$$\tau_{k+1} = \max_{\mathbf{r}_{0,k}, \bar{\mathbf{r}}_{0,k} \in \mathbb{M}^{k+1}} \left\{\|\mathbf{H}(\mathbf{r}_{0,k}) - \mathbf{H}(\bar{\mathbf{r}}_{0,k})\|^2\right\}$$

$$c_{k+1} = \max_{\mathbf{r}_{0,k} \in \mathbb{M}^{k+1}} \left\{\|\mathcal{F}(\mathbf{r}_{0,k})\|^2\right\}$$

$$\eta_{k+1} = \max_{\mathbf{r}_{0,k} \in \mathbb{M}^{k+1}} \left\{\|\mathcal{B}(\mathbf{r}_{0,k})\|^2\right\}$$

$$\delta_{k+1} = \min_{\mathbf{r} \in \mathbb{M}^{k+1}} \left\{\|\mathbf{H}(\mathbf{r}_{0,k})\|^2\right\}$$

$$\varphi_{k+1} = \max_{\mathbf{r}_{0,k} \in \mathbb{M}^{k+1}} \bar{\lambda}\left\{\mathbf{G}(\mathbf{r}_{0,k})\mathbf{G}(\mathbf{r}_{0,k})'\right\}$$

$$u = \max_{r \in \mathbb{M}} \left\{\|A_r\|^2\right\}, \; b = \max_{r \in \mathbb{M}} \left\{\|C_r\|^2\right\}$$

where $k \in [1, N]$, and $\mathcal{B}(r_{0,\alpha-1}) = [\mathcal{F}(\mathbf{r}_{0,\alpha-1})\, \mathcal{F}(\mathbf{r}_{1,\alpha-1}) \cdots I_{n_x \times n_x}]$. Furthermore, we assume that system Σ satisfies the following assumption.

Assumption 5.1: First, the MJLS Σ is stochastic observable. For all $\mathbf{r} \in \mathbb{M}^{N+1}$ with $N \geq \bar{\mathrm{t}}_0$, the corresponding observability matrix satisfies $\mathrm{rank}(\mathbf{H}(\mathbf{r})) = n_x$; and second, System Σ is (α, β)-mode stochastic observable in $N+1$ steps, and $N \geq \alpha + \beta + 1$.

Then, three virtual noise variables are defined to account for the possible mismatch between the true state $\mathbf{r}_{t-N,t}$ and its estimate $\hat{\mathbf{r}}_{t-N,t}$ within $[t-N, t]$. They are given as

$$\varepsilon_1 = \left[\mathbf{H}(\mathbf{r}_{t-N,t}^{\alpha,\beta}) - \mathbf{H}(\hat{\mathbf{r}}_{t-N,t}^{\alpha,\beta})\right] x_{t-N+\alpha}$$

$$\varepsilon_2 = \left[\mathbf{H}(\mathbf{r}_{t-N,t}^{\alpha,0}) - \mathbf{H}(\hat{\mathbf{r}}_{t-N,t}^{\alpha,0})\right] x_{t-N+\alpha}$$

$$\epsilon = \left[\mathbf{H}(\mathbf{r}_{t-N,t}^{N-\beta,0}) - \mathbf{H}(\hat{\mathbf{r}}_{t-N,t}^{N-\beta,0})\right] x_{t-\beta}$$

$$\chi_\alpha = [\mathcal{F}(\mathbf{r}_{t-N,t-N+\alpha-1}) - \mathcal{F}(\hat{\mathbf{r}}_{t-N,t-N+\alpha-1})]x_{t-N}$$

where $\varepsilon \in \mathbb{R}^{(N_r+1)n_y \times 1}$, $\epsilon \in \mathbb{R}^{(\beta+1)n_y \times 1}$, $\chi \in \mathbb{R}^{n_x \times 1}$ with $N_r = N - \alpha - \beta$; $\mathbf{r}_{t-N,t}^{\alpha,\beta}, \mathbf{r}_{t-N,t}^{N-\beta,0}, \mathbf{r}_{t-N,t-k} \subseteq \mathbf{r}$; $\hat{\mathbf{r}}_{t-N,t}^{\alpha,\beta}, \hat{\mathbf{r}}_{t-N,t}^{N-\beta,0}, \hat{\mathbf{r}}_{t-N,t-k} \subseteq \hat{\mathbf{r}}$; $\hat{\mathbf{r}}_{t-N,t}^{\alpha,\beta} \neq \mathbf{r}_{t-N,t}^{\alpha,\beta}$; $\hat{\mathbf{r}}_{t-N,t}^{N-\beta,0} \neq \mathbf{r}_{t-N,t}^{N-\beta,0}$; and $\hat{\mathbf{r}}_{t-N,t-k} \neq \mathbf{r}_{t-N,t-k}$.

We first introduce the following lemma as a supplement proof condition.

Lemma 5.1: Suppose that system Σ is (α, β)-mode observable in $N+1$ steps. There exists positive constants ρ_{ε_s}, ρ_ϵ, and ρ_{χ_k} with $s \in \{1, 2\}$ and $k \in \{1, 2, \ldots, N\}$ such that the expected value of state satisfies following inequality:

$$\mathrm{E}\left\{\|\varepsilon_s\|^2 \middle| x_{t-N}, \mathbf{r}_{t-N,t}\right\} \leq \rho_{\varepsilon_s}$$

$$\mathrm{E}\left\{\|\epsilon\|^2 \middle| x_{t-N}, \mathbf{r}_{t-N,t}\right\} \leq \rho_\epsilon$$

$$\mathrm{E}\left\{\|\chi_k\|^2 \middle| x_{t-N}, \mathbf{r}_{t-N,t}\right\} \leq \rho_{\chi_k}.$$

See Appendix C for the proof. ■

A. Main Results

Inspired by the method given in [20], the argumentation is given under a stochastic environment based on seeking the upper and lower bounds on the expected value of cost function J.

Then, we give the following three theorems on the stability of the MD-MHE scheme algorithms.

1) Stability of Mode Detection-Moving Horizon Estimator:

Theorem 5.1: Suppose that Assumption 5.1 holds. The mean square norm of the estimation error of the MD-MHE estimator is bounded as

$$\mathrm{E}\left\{\|\hat{e}_{t-N+\alpha}\|^2|x_{t-N},\mathbf{r}_{t-N,t}\right\}\leq\xi_t$$

where $\{\xi_t\}$ is a sequence generated by

$$\xi_t=\frac{g_0}{f_0}\xi_{t-1}+\frac{d_0}{f_0},\ \xi_\alpha=\bar{d}_0$$

with

$$\begin{aligned}
f_0&=f=\frac{1}{2}\underline{\lambda}(P)+\frac{l\delta_{N_r+1}\underline{\lambda}(R)}{3(l+1)}\\
l&=(\underline{\lambda}(Q)-2\varphi_{N_r+1}\bar{\lambda}(R))\\
g_0&=6\bar{\lambda}(P)\left(u+\frac{\bar{\lambda}(R)}{\underline{\lambda}(Q)-4bu\bar{\lambda}(R)}\right)\\
d_0&=6\bar{\lambda}(P)\left(\rho_{\chi_1}+\frac{4gb\rho_{\chi_1}\bar{\lambda}(R)}{\underline{\lambda}(Q)-4bu\bar{\lambda}(R)}\right.\\
&\quad\left.+\left(\frac{\bar{\lambda}(Q)+4bu\bar{\lambda}(R)}{\underline{\lambda}(Q)-4bu\bar{\lambda}(R)}+\frac{1}{3}\right)\mu_w\right)+d\\
d&=\frac{l\rho_{\varepsilon_1}}{l+1}+\varphi_{N_r+1}\bar{\lambda}(R)\frac{N_r\left(\bar{\lambda}(Q)+2\varphi_{N_r+1}\bar{\lambda}(R)\right)\mu_w}{l+1}\\
&\quad+\frac{lN_r\varphi_{N_r+1}}{l+1}\bar{\lambda}(R)\mu_w+\frac{l(N_r+1)}{l+1}\bar{\lambda}(R)\mu_v\\
&\quad+N_r\bar{\lambda}(Q)\mu_w+(N_r+1)\bar{\lambda}(R)\mu_v\\
\bar{d}_0&=\frac{g}{f}\left\{2\rho_{\chi_\alpha}+2c_\alpha\bar{\lambda}(P_0)+\alpha\eta_\alpha\mu_w\right\}+\frac{d}{f},\ g=2\bar{\lambda}(P).
\end{aligned}$$

The proof is given in Appendix D. ■

Remark 5.1: Note that if $0<\frac{g_0}{f_0}<1$, the bounding sequence $\{\xi_t\}$ converges exponentially to the asymptotic value

$$\xi_\infty=\frac{d_0}{f_0-g_0}.$$

We notice that, for any $u>0$ and $\delta>0$, the condition $0<\frac{g_0}{f_0}<1$ can be easily satisfied by selecting appropriate values for P, Q, and R.

2) Stability of Extended Mode Detection-Moving Horizon Estimator:

Theorem 5.2: Suppose that Assumption 5.1 holds. The mean square norm of the estimation error of the EMD-MHE estimator is bounded as

$$\mathrm{E}\left\{\|\hat{e}^{(e)}_{t-N+\alpha}\|^2|x_{t-N},\mathbf{r}_{t-N,t}\right\}\leq\xi^{(e)}_t$$

where $\{\xi^{(e)}_t\}$ is a sequence generated by

$$\xi^{(e)}_t=\frac{g^{(e)}_0}{f^{(e)}_0}\xi^{(e)}_{t-1}+\frac{d^{(e)}_0}{f^{(e)}_0},\ \xi^{(e)}_\alpha=\bar{d}^{(e)}_0$$

with

$$\begin{aligned}
f^{(e)}_0&=\frac{1}{2}\underline{\lambda}(P)+\frac{l^{(e)}\delta_{N-\alpha+1}\zeta\underline{\lambda}(R)}{3(l^{(e)}+1)}\\
l^{(e)}&=\underline{\lambda}(Q)-2\varphi_{N-\alpha+1}\bar{\lambda}(R)\\
g^{(e)}_0&=6\bar{\lambda}(P)\left(u+\frac{\bar{\lambda}(R)}{\underline{\lambda}(Q)-4bu\bar{\lambda}(R)}\right)\\
d^{(e)}_0&=d^{(e)}_1+d^{(e)}_2+d^{(e)}_3\\
d^{(e)}_1&=\frac{l^{(e)}\rho_{\varepsilon_2}}{l^{(e)}+1}+\varphi_{N-\alpha+1}\bar{\lambda}(R)\frac{(N-\alpha)\left[\bar{\lambda}(Q)+2\varphi_{N-\alpha+1}\mu_w\right]}{l^{(e)}+1}\\
&\quad+\frac{l^{(e)}(N-\alpha)\varphi_{N-\alpha+1}}{l^{(e)}+1}\bar{\lambda}(R)\mu_w+\frac{l^{(e)}(N_r+1)}{l^{(e)}+1}\bar{\lambda}(R)\mu_v\\
&\quad+\frac{l^{(e)}}{l^{(e)}+1}\zeta\beta\bar{\lambda}(R)\mu_v\\
d^{(e)}_2&=\bar{\lambda}(Q)(N-\alpha)\mu_w+(N_r+1)\bar{\lambda}(R)\mu_v+\zeta\beta\bar{\lambda}(R)\mu_v\\
d^{(e)}_3&=6\bar{\lambda}(P)\left(\rho_{\chi_1}+\frac{4gb\rho_{\chi_1}\bar{\lambda}(R)}{\underline{\lambda}(Q)-4bu\bar{\lambda}(R)}+\left(\frac{\bar{\lambda}(Q)+4bu\bar{\lambda}(R)}{\underline{\lambda}(Q)-4bu\bar{\lambda}(R)}\right.\right.\\
&\quad\left.\left.+\frac{1}{3}\right)\mu_w\right)\\
\bar{d}^{(e)}_0&=\frac{g}{f^{(e)}_0}\left\{2\rho_{\chi_\alpha}+2c_\alpha\bar{\lambda}(P_0)+\alpha\eta_\alpha\mu_w\right\}\\
&\quad+\frac{d^{(e)}_1+d^{(e)}_2}{f^{(e)}_0}.
\end{aligned}$$

The proof is given in Appendix E. ■

Remark 5.2: If $0<\frac{g^{(e)}_0}{f^{(e)}_0}<1$, the bounding sequence $\{\xi^{(e)}_t\}$ converges exponentially to the asymptotic value

$$\xi^{(e)}_\infty=\frac{d^{(e)}_0}{f^{(e)}_0-g^{(e)}_0}.$$

Using the same principle, the design parameters P, Q, R, and ζ can be easily chosen to satisfy the convergent condition $\frac{g^{(e)}_0}{f^{(e)}_0}<1$ for any $u>0$ and $\delta>0$.

3) Stability of Incomplete Mode Detection-Moving Horizon Estimator:

Theorem 5.3: Suppose that Assumption 5.1 holds. The mean square norm of the estimation error of the IMD-MHE estimator is bounded as

$$\mathrm{E}\left\{\|\hat{e}^{(i)}_{t-N+\alpha}\|^2|x_{t-N},\mathbf{r}_{t-N,t}\right\}\leq\xi^{(i)}_t$$

where $\{\xi^{(i)}_t\}$ is a sequence generated by

$$\xi^{(i)}_t=\frac{g^{(i)}_0}{f^{(i)}_0}\xi^{(i)}_{t-1}+\frac{d^{(i)}_0}{f^{(i)}_0},\ \xi^{(i)}_\alpha=\bar{d}^{(i)}_0$$

with

$$f_0^{(i)} = \frac{1}{2}\underline{\lambda}(P) + \frac{l^{(i)}\delta_{\beta+1}\zeta\underline{\lambda}(R)}{3(l^{(i)}+1)}$$

$$l^{(i)} = \underline{\lambda}(Q) - 2\varphi_{\beta+1}\bar{\lambda}(R)$$

$$g_0^{(i)} = 6\bar{\lambda}(P)\left(u + \frac{\bar{\lambda}(R)}{\underline{\lambda}(Q) - 4bu\bar{\lambda}(R)}\right)$$

$$d_0^{(i)} = d_1^{(i)} + d_2^{(i)} + d_3^{(i)}$$

$$d_1^{(i)} = \frac{l^{(i)}\rho_\epsilon}{l^{(i)}+1} + \varphi_{\beta+1}\bar{\lambda}(R)\frac{\beta\left[\bar{\lambda}(Q) + 2\varphi_{\beta+1}\mu_w\right]}{l^{(i)}+1} + \frac{l^{(i)}\beta\varphi_{\beta+1}}{l^{(i)}+1}\bar{\lambda}(R)\mu_w + \frac{l^{(i)}(\beta+1)}{l^{(i)}+1}\zeta\bar{\lambda}(R)\mu_v$$

$$d_2^{(i)} = \bar{\lambda}(Q)\beta\mu_w + \zeta(\beta+1)\bar{\lambda}(R)\mu_v$$

$$d_3^{(i)} = 6\bar{\lambda}(P)\left(\rho_{\chi_1} + \frac{4gb\rho_{\chi_1}\bar{\lambda}(R)}{\underline{\lambda}(Q) - 4bu\bar{\lambda}(R)} + \left(\frac{\bar{\lambda}(Q) + 4bu\bar{\lambda}(R)}{\underline{\lambda}(Q) - 4bu\bar{\lambda}(R)} + \frac{1}{3}\right)\mu_w\right)$$

$$\bar{d}_0^{(i)} = \frac{g}{f_0^{(i)}}\left\{2\rho_{\chi_{N-\beta}} + 2c_{N-\beta}\bar{\lambda}(P_0) + (N-\beta)\eta_{N-\beta}\mu_w\right\} + \frac{d_1^{(i)} + d_2^{(i)}}{f_0^{(i)}}.$$

The proof is given in Appendix F. ■

Remark 5.3: As in Theorem 5.2, if P, Q, R, and ζ are selected such that $0 < \frac{g_0^{(i)}}{f_0^{(i)}} < 1$, the bounding sequence $\{\xi_t^{(i)}\}$ converges exponentially to the asymptotic value

$$\xi_\infty^{(i)} = \frac{d_0^{(i)}}{f_0^{(i)} - g_0^{(i)}}.$$

VI. Verification Simulation

The filters described by the three MD-MHE-based algorithms are applied to a target tracking problem. We compare their performances with the performance of Kalman filter (KF) algorithm with perfect information of the mode jumping process. In addition, detailed comparison of the performances is given between the new estimator algorithms and the IMM algorithm, which is a standard "classical" multimodel filtering algorithm based on a bank of filters.

Consider the following oscillator modeled as a discrete MJLS with two modes over oscillation frequencies. This model is a modified version of that given in [27]

$$A(1) = \begin{bmatrix} \cos(w_1T_s) & -w_1\sin(w_1T_s) \\ \frac{1}{w_1}\sin(w_1T_s) & \cos(w_1T_s) \end{bmatrix}$$

$$A(2) = \begin{bmatrix} \cos(w_2T_s) & -w_2\sin(w_2T_s) \\ \frac{1}{w_2}\sin(w_2T_s) & \cos(w_2T_s) \end{bmatrix}$$

$$C(1) = C(2) = [1,\ 0].$$

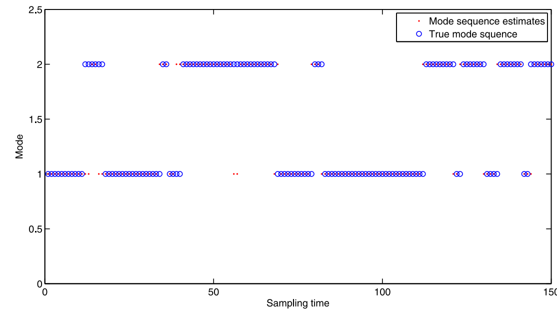


Fig. 2. Tracking of the mode jumping process showing displacement against time.

where state $x \triangleq [x(1), x(2)]'$, the mode parameters $w_1 = \frac{1}{2}$, $w_2 = 1$, and the sampling time $T_s = 0.5$. We assume that the jumping process between the modes is governed by a Markov chain taking values form $\{1, 2\}$, and the transition matrix is given by

$$\Pi = \begin{bmatrix} 0.925 & 0.075 \\ 0.075 & 0.925 \end{bmatrix}. \tag{18}$$

The initial state is independent, and with Gaussian distribution $\mathcal{N}(x_0; 10, 1)$. The process noise and measurement noise are independent and Gaussian distributed with zero mean, and covariances $P_w = [1,\ 0;\ 0,\ 1]$ and $P_v = 2$. The corresponding parameters $\mu_w = \mu_v = 2$.

By analyzing the (α, β)-mode stochastic observation for the MJLS in Definition 3.4, the horizon size is chosen to be $N = 13$ with $\alpha = 3$ and $\beta = 4$. For the design of the MD-MHE scheme estimators, the designed parameters are chosen to be $P = 1$, $Q = 0.1$, $R = 0.04$, and $\zeta^{-1} = 1.2$, which satisfy Assumption 5.1 and Theorems 5.1–5.3.

All the designed estimators, KF, and IMM filters are simulated with $T = 150$ sampling times for a 25-time Monte Carlo simulation using MATLAB. The performance measure is the root mean square (RMS) error of $x(i)$ computed as

$$\text{RMS} = \sqrt{\frac{1}{T}\sum\nolimits_{t=1}^{T}\left(x_t^{(s)}(i) - \hat{x}_t^{(s)}(i)\right)^2} \tag{19}$$

where $\hat{x}_t^{(s)}(i)$ is the state estimates of the elements $x(i)$ in the x direction at time t of the sth Monte Carlo simulation. Fig. 2 shows the tracking result of the mode jumping process under the mode detection algorithm based on the ML algorithm. The result shows that for the estimated track of the mode jump sequence, the ML-based mode detection algorithm can track the mode switching process of the MJLS. The estimated tracks obtained by the proposed MHE-based estimators, and IMM and KF are shown in Figs. 3 and 4. Fig. 5 shows their average estimation RMS errors for the tracking of $x(1)$, suggesting that all the three MHE related algorithms outperform IMM in terms of estimation errors. On the other hand, the utilization of the reliable factor ζ, the EMD-MHE algorithm, and IMD-MHE algorithm can provide the state estimates without any time delay compared with the MD-MHE algorithm. The average RMS errors of the five estimators obtained from one 150-times simulation for $x(1)$ and $x(2)$ are presented in Table I. It shows that the average RMS errors of MD-MHE and EMD-MHE are close to the estimated

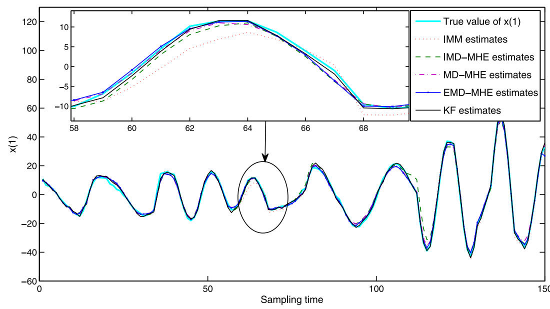


Fig. 3. Tracking of state $x(1)$ showing displacement against time.

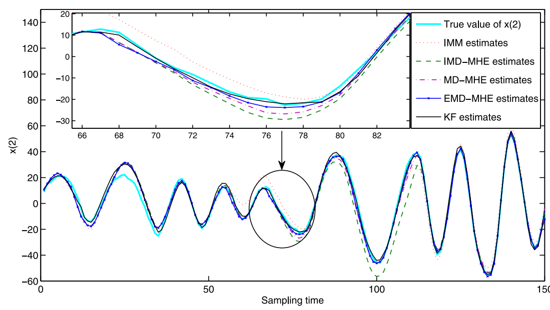


Fig. 4. Tracking of state $x(2)$ showing displacement against time.

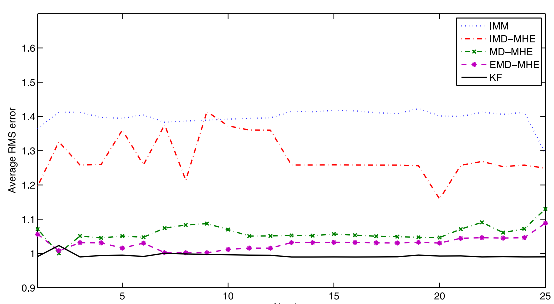


Fig. 5. Comparison of the average RMS errors of state estimate of $x(1)$.

TABLE I
AVERAGE RMS ERRORS AND COMPUTATIONAL COST FOR THE FIVE ESTIMATOR ALGORITHMS

Algorithm	KF	MD-MHE	EMD-MHE
Average RMS error of $x(1)$	0.9899	1.0511	1.0388
Average RMS error of $x(2)$	2.5760	2.9834	2.7751
Computational time (s)	0.018	1.255	2.361
Algorithm	IMD-MHE	IMM	
Average RMS error of $x(1)$	1.2636	1.5310	
Average RMS error of $x(2)$	3.8484	4.4533	
Computational time (s)	1.006	0.100	

track of KF that has exact knowledge of the mode jumping process.

Remark 6.1: Clearly, there exist advantages of the three MD-based MHE algorithms, as evident from the RMS errors and computational cost given in Table I. We consider the following two aspects.

1) Since more available observable information is utilized in the EMD-MHE algorithm, state estimate can be more accurately obtained than other two algorithms.
2) With a shorter moving horizon, the IMD-MHE algorithm can provide the solutions with less computational complexity.

Consider the computational time of each estimator in Table I. The three MHE related algorithms have a computational cost 10 times higher than IMM. The cause of this deficiency has two aspects. First, the computational complexity of solving a quadratic program means that the time required for the MHE-based algorithm schemes is longer than those for IMM and KF algorithms. Second, a larger amount of stored data is required for the MHE-based algorithms than for the IMM and KF algorithms. Since the horizon length of the EMD-MHE is longer than that of the MD-MHE and IMD-MHE algorithms, it requires higher data storage, and takes more than 2 s to obtain the solution of the quadratic program at each step. All computations were performed on a 2.6 GHz Intel processor with 4 GB installed memory. Hence, it still has room for improvement with regard to these two aspects for the proposed estimator design algorithms.

VII. CONCLUSION

We proposed a new optimization-based state estimation algorithm for discrete-time MJLSs. The maximum-likelihood approach was introduced and combined with the moving horizon estimation method to solve the mode detection-based moving horizon state estimation problem. State estimation algorithms were designed using the mode stochastic observable results of MJLSs, they are the MD-MHE, the EMD-MHE, and the incomplete MD-MHE. Based on the observability analysis of the discrete-time MJLSs, we provided a set of sufficient conditions for the stability of the three new estimator algorithms. The effectiveness of the proposed algorithms was verified with a numerical example.

APPENDIX

A. Proof of Theorem 3.1

Proof: In view of (12), we obtain the optimal estimates $x_0^{ML}(\mathbf{r})$ in the maximum-likelihood sense conditioned on $\mathbf{r}$ by solving the following optimization problem:

$$\begin{aligned} x_0^{ML}(\mathbf{r}) &= \arg \max_{x_0 \in \mathbb{R}^{n_x}} \log \{ f(\mathbf{y}_{0,N}, x_0 | \mathbf{r}) \} \\ &= \arg \min_{x_0 \in \mathbb{R}^{n_x}} d(\mathbf{y}_{0,N}, \mathbf{r}, x_0) + \log |\Delta(\mathbf{r})|. \end{aligned} \tag{20}$$

By evaluating

$$\left. \frac{\partial d(\mathbf{y}_{0,N}, \mathbf{r}, x_0)}{\partial x_0} \right|_{x_0 = x_0^{ML}(\mathbf{r})} = 0$$

$x_0^{ML}(\mathbf{r})$ can be computed from

$$x_0^{ML}(\mathbf{r}) = [\mathbf{H}(\mathbf{r})' \Delta(\mathbf{r})^{-1} \mathbf{H}(\mathbf{r})]^{-1} \mathbf{H}(\mathbf{r})' \Delta(\mathbf{r})^{-1} \mathbf{y}_{0,N}.$$

Denoting the minimum distance $\bar{d}(\mathbf{y}_{0,N},\mathbf{r})$ by

$$\bar{d}(\mathbf{y}_{0,N},\mathbf{r}) \triangleq \min_{x_0\in\mathbb{X}} d(\mathbf{y}_{0,N},\mathbf{r},x_0) = \left\| \left[I-\tilde{P}(\mathbf{r})\right]\mathbf{y}_{0,N}\right\|^2_{\Delta(\mathbf{r})^{-1}}$$

where

$$\tilde{P}(\mathbf{r}) = \mathbf{H}(\mathbf{r})[\mathbf{H}(\mathbf{r})'\Delta(\mathbf{r})^{-1}\mathbf{H}(\mathbf{r})]^{-1}\mathbf{H}(\mathbf{r})'\Delta(\mathbf{r})^{-1}.$$

Then

$$\max_{x_0\in\mathbb{R}^{n_x}} (\log f(\mathbf{y}_{0,N},x_0|\mathbf{r})) = (\log f(\mathbf{y}_{0,N},x_0|\mathbf{r}))\Big|_{x_0^{ML}(\mathbf{r})} = -\frac{1}{2}\{\log\{|\Delta(\mathbf{r})|\}+\bar{d}(\mathbf{y}_{0,N},\mathbf{r})\} - \frac{1}{2}\log(2\pi). \tag{21}$$

Combining (11) and (21) yields

$$\hat{\mathbf{r}}_{0,N} = \arg\min_{\mathbf{r}\in\mathbb{M}^{N+1}} \big(-2\log\{\Pr(\mathbf{r})\}+\log\{|\Delta(\mathbf{r})|\} + \bar{d}(\mathbf{y}_{0,N},\mathbf{r})\big)$$

where we neglect the additive constants that have little impact on the maximization problem. Thus, for any $\mathbf{r}\in\mathbb{M}^{N+1}$

$$\mathbf{Pr}(\hat{\mathbf{r}}_{0,N}\succ\mathbf{r}|x_0,\tilde{\mathbf{r}}) = \Big\{-2\log\{\Pr(\mathbf{r})\}+\log\{|\Delta(\mathbf{r})|\}+\bar{d}(\mathbf{y}_{0,N},\mathbf{r}) < -2\log\{\Pr(\hat{\mathbf{r}}_{0,N})\}+\log|\Delta(\hat{\mathbf{r}}_{0,N})|+\bar{d}(\mathbf{y}_{0,N},\hat{\mathbf{r}}_{0,N})|x_0,\tilde{\mathbf{r}}\Big\}. \tag{22}$$

This expression is the desired result.

B. Proof of Theorem 3.2

Proof: The proof is similarly to the technique used in [27], and is not shown for brevity. ■

C. Proof of Lemma 5.1

Proof: Given x_{t-N} and $\mathbf{r}_{t-N,t}$, We consider the expected value of $\|\varepsilon_1\|^2$ for the following two cases.

Case 1: If $\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t} = \mathbf{r}^{\alpha,\beta}_{t-N,t}$, then $\varepsilon_1 = \mathbf{0}_{(N_r+1)n_y\times 1}$, where $N_r = N-\alpha-\beta$.

Case 2: if $\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t} \neq \mathbf{r}^{\alpha,\beta}_{t-N,t}$, the value of $\|\varepsilon_1\|^2$ satisfies the following inequality:

$$\|\varepsilon_1\|^2 = \left\|\left[\mathbf{H}(\mathbf{r}^{\alpha,\beta}_{t-N,t}) - \mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\right]x_{t-N+\alpha}\right\|^2_{\tilde{\mathcal{R}}} \le \bar{\lambda}(R)\tau_{N_r+1}c_{\alpha-1}\|x_{t-N}\|^2.$$

It follows that

$$\mathrm{E}\{\|\varepsilon_1\|^2|x_{t-N},\mathbf{r}_{t-N,t}\} \le \sum_{\mathbf{s}^{\alpha,\beta}_{t-N,t}\neq\mathbf{r}^{\alpha,\beta}_{t-N,t}} \mathbf{Pr}\left\{\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t} = \mathbf{s}^{\alpha,\beta}_{t-N,t}|x_{t-N},\mathbf{r}_{t-N,t}\right\} \times\bar{\lambda}(R)\tau_{N_r+1}c_{\alpha-1}\|x_{t-N+\alpha}\|^2 \tag{23}$$

where $\mathbf{s}_{t-N,t}\in\mathbb{M}^{N+1}$. It can be shown that the probability $\mathbf{Pr}\left\{\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t} = \mathbf{s}^{\alpha,\beta}_{t-N,t}|x_{t-N},\mathbf{r}_{t-N,t}\right\}$ satisfies the following inequality:

$$\mathbf{Pr}\left\{\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t} = \mathbf{s}^{\alpha,\beta}_{t-N,t}|x_{t-N},\mathbf{r}_{t-N,t}\right\} = \sum_{\mathbf{s}^{\alpha,\beta}_{t-N,t}=\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t}} \mathbf{Pr}\left\{\mathbf{s}_{t-N,t}|x_{t-N},\mathbf{r}_{t-N,t}\right\} \le \sum_{\mathbf{s}^{\alpha,\beta}_{t-N,t}=\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t}} \Big(1-\mathbf{Pr}\left\{\mathbf{r}_{t-N,t}\succ\mathbf{s}_{t-N,t}|x_{t-N},\mathbf{r}_{t-N,t}\right\}\Big). \tag{24}$$

Using Theorem 3.2, the inequality can be rewritten as

$$\mathbf{Pr}\left\{\hat{\mathbf{r}}_{t-N,t} = \mathbf{s}^{\alpha,\beta}_{t-N,t}|x_{t-N},\mathbf{r}_{t-N,t}\right\} \le \sum_{s^{\alpha,\beta}_{t-N,t}=\hat{\mathbf{r}}^{\alpha,\beta}_t} \left(1-\gamma_{(Nn_y)}\left(\hbar+\vartheta\|x_{t-N}\|^2\right)\right)$$

where

$$\hbar = \min_{\mathbf{r}_{t-N,t},\mathbf{s}_{t-N,t}\in\mathbb{M}^{N+1}} \frac{2}{\varsigma}\log\frac{\Pr(\mathbf{r}_{t-N,t})}{\Pr(\mathbf{s}_{t-N,t})} + \frac{1}{\varsigma}\log\frac{|\Delta(\mathbf{s}_{t-N,t})|}{|\Delta(\mathbf{r}_{t-N,t})|}$$

$$\vartheta = \min_{\mathbf{r}_{t-N,t},\mathbf{s}_{t-N,t}\in\mathbb{M}^{N+1}} \frac{\kappa}{\varsigma}.$$

The expected value of (23) will be upper bounded by

$$\mathrm{E}\{\|\varepsilon_1\|^2|x_{t-N},\mathbf{r}_{t-N,t}\} \le \mathrm{Card}(\mathbb{A})\bar{\lambda}(R)\tau_{N_r+1}c_{\alpha-1}\|x_{t-N}\|^2\Big(1-\gamma_{(N+1)n_y} \times\left(\hbar+\vartheta\|x_{t-N}\|^2\right)\Big)$$

where $\mathrm{Card}(\mathbb{A})$ is the cardinality of a set $\mathbb{A}$, $\mathbb{A}\subseteq\mathbb{M}^{N+1}$, and $\gamma_{(N+1)n_y}(\cdot)$ is the cumulative distribution function of a χ^2 random variable with $(N+1)n_y$ degrees of freedom. It follows that

$$1-\gamma_{(N+1)n_y}(x) = \mathbf{\Gamma}\left(\frac{(N+1)n_y}{2},\frac{x}{2}\right)$$

where $\mathbf{\Gamma}(\cdot,\cdot)$ is a regularized Gamma function. The function has a super exponential convergence to 0 as $x\to+\infty$. Therefore, we have the following equality:

$$\lim_{\|x_{t-N}\|^2\to+\infty} \mathrm{Card}(\mathbb{A})\bar{\lambda}(R)\tau_{N_r+1}c_{\alpha-1}\|x_{t-N}\|^2 \times\mathbf{\Gamma}\left(\frac{(N+1)n_y}{2},\frac{\hbar+\vartheta\|x_{t-N}\|^2}{2}\right) = 0.$$

We can infer from Cases 1 and 2 that the upper bound depends continuously on $\mathrm{E}\left\{\|\varepsilon_1\|^2|x_{t-N},\mathbf{r}_{t-N,t}\right\}\in[0,+\infty)$, and there must exist a positive upper bound ρ_{ε_1}, which is independent of x_{t-N} and $\mathbf{r}_{t-N,t}$. Using the same approach, the upper bounds of $\mathrm{E}\left\{\|\varepsilon_2\|^2|x_{t-N},\mathbf{r}_{t-N,t}\right\}$, $\mathrm{E}\left\{\|\epsilon\|^2|x_{t-N},\mathbf{r}_{t-N,t}\right\}$, and $\mathrm{E}\left\{\|\chi_k\|^2|x_{t-N},\mathbf{r}_{t-N,t}\right\}$ can be obtained and called ρ_{ε_2}, ρ_ϵ, and ρ_{χ_k}, respectively.

D. Proof of Theorem 5.1

Proof: The optimal cost J^* is given as

$$\begin{aligned}&J^*(\hat{x}_{t-N+\alpha},\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t},\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t})\\&\quad=\|\hat{x}_{t-N+\alpha}-\bar{x}_{t-N+\alpha}\|^2_P+\|\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{Q}}}\\&\qquad+\|\mathbf{y}^{t-\beta}_{t-N+\alpha}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{x}_{t-N+\alpha}-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}.\end{aligned}\tag{25}$$

From the optimal solution of $\{\hat{x}_{t-N+\alpha},\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\}$ for the cost function, we have

$$\begin{aligned}&J^*(\hat{x}_{t-N+\alpha},\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t},\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t})\\&\quad\le\|x_{t-N+\alpha}-\bar{x}_{t-N+\alpha}\|^2_P+\|\mathbf{w}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{Q}}}+\|\mathbf{v}^{\alpha,\beta}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}.\end{aligned}$$

Using the linearity of expectation and Proposition 2.1, the expected value of the cost function is

$$\begin{aligned}&\mathrm{E}\left\{J^*(\hat{x}_{t-N+\alpha},\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t},\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t})\right\}\\&\quad\le\mathrm{E}\left\{\|x_{t-N+\alpha}-\bar{x}_{t-N+\alpha}\|^2_P|\cdot\right\}+E\left\{\|\mathbf{w}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{Q}}}\right\}\\&\qquad+\mathrm{E}\left\{\|\mathbf{v}^{\alpha,\beta}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}\right\}\\&\quad\le\mathrm{E}\left\{\|x_{t-N+\alpha}-\bar{x}_{t-N+\alpha}\|^2_P|\cdot\right\}+N_r\overline{\lambda}(Q)\mu_w\\&\qquad+(N_r+1)\overline{\lambda}(R)\mu_v\end{aligned}\tag{26}$$

where $\mathrm{E}\{\|\cdot\||\cdot\}\triangleq\mathrm{E}\{\|\cdot\||x_{t-N},\mathbf{r}_{t-N,t}\}$. To find a lower bound on the optimal cost J^* in terms of the lower bounded of the first term on the right-hand side of (25), the following inequality can be obtained by using the property of the quadratic function:[1]:

$$\begin{aligned}&\mathrm{E}\left\{\|x_{t-N+\alpha}-\hat{x}_{t-N+\alpha}\|^2|\cdot\right\}\\&\quad\le\mathrm{E}\left\{\left\|x_{t-N+\alpha}-\bar{x}_{t-N+\alpha}+\bar{x}_{t-N+\alpha}-\hat{x}_{t-N+\alpha}\right\|^2|\cdot\right\}\\&\quad\le2\mathrm{E}\left\{\|x_{t-N+\alpha}-\bar{x}_{t-N+\alpha}\|^2|\cdot\right\}\\&\qquad+2\mathrm{E}\left\{\|\bar{x}_{t-N+\alpha}-\hat{x}_{t-N+\alpha}\|^2|\cdot\right\}.\end{aligned}\tag{27}$$

[1]Let $x_1,x_2,\ldots,x_q\in\mathbb{R}^n$. Then, the following inequalities hold [20]

$$\left\|\sum_{i=1}^q x_i\right\|^2\le q\sum_{i=1}^q\|x_i\|^2.$$

Rewrite (27) as

$$\begin{aligned}&\mathrm{E}\left\{\|\bar{x}_{t-N+\alpha}-\hat{x}_{t-N+\alpha}\|^2|\cdot\right\}\\&\quad\ge\frac{1}{2}\mathrm{E}\left\{\|x_{t-N+\alpha}-\hat{x}_{t-N+\alpha}\|^2|\cdot\right\}\\&\qquad-\mathrm{E}\left\{\|x_{t-N+\alpha}-\bar{x}_{t-N+\alpha}\|^2|\cdot\right\}.\end{aligned}\tag{28}$$

To obtain the lower bound of the second term on the right-hand side of (25), we use the following expression:

$$\begin{aligned}&\mathrm{E}\left\{\left\|\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\left(x_{t-N+\alpha}-\hat{x}_{t-N+\alpha}\right)\right\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\right\}\\&\quad=\mathrm{E}\Big\{\Big\|\mathbf{y}^{t-\beta}_{t-N+\alpha}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{x}_{t-N+\alpha}-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\\&\qquad+\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}+\left(\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})-\mathbf{H}(\mathbf{r}^{\alpha,\beta}_{t-N,t})\right)x_{t-N+\alpha}\\&\qquad-\mathbf{G}(\mathbf{r}^{\alpha,\beta}_{t-N,t})\mathbf{w}^{\alpha,\beta+1}_{t-N,t}-\mathbf{v}^{\alpha,\beta}_{t-N,t}\Big\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\\&\quad\le\mathrm{E}\Big\{3\|\mathbf{y}^{\alpha,\beta}_{t-N,t}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{x}_{t-N+\alpha}\\&\qquad-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\\&\qquad+\mathrm{E}\Big\{3\|\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}\Big\}\\&\qquad+\mathrm{E}\Big\{3\Big\|\left(\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})-\mathbf{H}(\mathbf{r}^{\alpha,\beta}_{t-N,t})\right)x_{t-N+\alpha}\Big\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\\&\qquad+\mathrm{E}\Big\{3\|\mathbf{G}(\mathbf{r}^{\alpha,\beta}_{t-N,t})\mathbf{w}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}\Big\}+\mathrm{E}\Big\{3\|\mathbf{v}^{\alpha,\beta}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}\Big\}.\end{aligned}$$

Then, we have

$$\begin{aligned}&\mathrm{E}\Big\{\Big\|\mathbf{y}^{\alpha,\beta}_{t-N,t}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{x}_{t-N+\alpha}-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\Big\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\\&\quad\ge\mathrm{E}\Big\{\frac{1}{3}\Big\|\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\left(x_{t-N+\alpha}-\hat{x}_{t-N+\alpha}\right)\Big\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\\&\qquad-\mathrm{E}\Big\{\Big\|\left(\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})-\mathbf{H}(\mathbf{r}^{\alpha,\beta}_{t-N,t})\right)x_{t-N+\alpha}\Big\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\\&\qquad-\mathrm{E}\Big\{\|\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}\Big\}\\&\qquad-\mathrm{E}\Big\{\|\mathbf{G}(\mathbf{r}^{\alpha,\beta}_{t-N,t})\mathbf{w}^{\alpha,\beta}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}\Big\}-\mathrm{E}\Big\{\|\mathbf{v}^{\alpha,\beta}_{t-N,t}\|^2_{\tilde{\mathcal{R}}}\Big\}.\end{aligned}\tag{29}$$

We now return to consider the second term $\mathrm{E}\Big\{\|\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\|\Big\}$, which satisfies

$$\begin{aligned}&J^*(\hat{x}_{t-N+\alpha},\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t},\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t})\\&\quad\le\|\hat{x}_{t-N+\alpha}-\bar{x}_{t-N+\alpha}\|^2_P+\left\|\mathbf{w}^{\alpha,\beta+1}_{t-N,t}\right\|^2_{\tilde{\mathcal{Q}}}\\&\qquad+\left\|\mathbf{y}^{t-\beta}_{t-N+\alpha}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{x}_{t-N+\alpha}-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\mathbf{w}^{\alpha,\beta+1}_{t-N,t}\right\|^2_{\tilde{\mathcal{R}}}.\end{aligned}\tag{30}$$

Rearranging (30) and computing its expected value, we obtain

$$\begin{aligned}&\mathrm{E}\left\{\|\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{Q}}}\right\}\\&\le \mathrm{E}\left\{\|\mathbf{w}^{\alpha,\beta+1}_{t-N,t}\|^2_{\tilde{\mathcal{Q}}}\right\}+\Big\|\mathbf{y}^{t-\beta}_{t-N+\alpha}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{x}_{t-N+\alpha}\\&\quad-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\Big\|^2_{\tilde{\mathcal{R}}}\\&\quad+2\mathrm{E}\left\{\left\|\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\left(\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}-\mathbf{w}^{\alpha,\beta+1}_{t-N,t}\right)\right\|^2_{\tilde{\mathcal{R}}}\right\}.\end{aligned}$$

Furthermore

$$\begin{aligned}&(\underline{\lambda}(Q)-2\varphi_{N_r+1}\bar{\lambda}(R))\mathrm{E}\left\{\|\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\|^2\right\}\\&\le N_r\bar{\lambda}(Q)\mu_w+2N_r\varphi_{N_r+1}\bar{\lambda}(R)\mu_w\\&\quad+\Big\|\mathbf{y}^{t-\beta}_{t-N+\alpha}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{x}_{t-N+\alpha}-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\Big\|^2_{\tilde{\mathcal{R}}}.\end{aligned}$$

where the parameters are set to satisfy $(\underline{\lambda}(Q)-2\varphi_{N_r+1}\bar{\lambda}(R))>0$. For brevity, let $l:=(\underline{\lambda}(Q)-2\varphi_{N_r+1}\bar{\lambda}(R))$. Then, we write (29) as

$$\begin{aligned}&\mathrm{E}\Big\{\Big\|\mathbf{y}^{\alpha,\beta}_{t-N,t}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{x}_{t-N+\alpha}-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N,t})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N,t}\Big\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\\&\ge\frac{l\delta_{N_r+1}\underline{\lambda}(R)}{3(l+1)}\mathrm{E}\Big\{\big\|\hat{e}_{t-N+\alpha}\big\|^2\big|\cdot\Big\}-\frac{l\rho_\varepsilon}{l+1}\\&\quad-\varphi_{N_r+1}\bar{\lambda}(R)\frac{N_r\left(\bar{\lambda}(Q)+2\varphi_{N_r+1}\bar{\lambda}(R)\right)\mu_w}{l+1}\\&\quad-\frac{lN_r\varphi_{N_r+1}}{l+1}\bar{\lambda}(R)\mu_w-\frac{l(N_r+1)}{l+1}\bar{\lambda}(R)\mu_v.\end{aligned}\tag{31}$$

Substituting (28) and (31) into (25), the lower bound of the expectation of the optimal cost J^* can be computed as

$$\begin{aligned}\mathrm{E}\{J^*\}\ge&\frac{1}{2}\underline{\lambda}(P)\mathrm{E}\Big\{\big\|\hat{e}_{t-N+\alpha}\big\|^2\big|\cdot\Big\}-\bar{\lambda}(P)\mathrm{E}\left\{\|\bar{e}_{t-N+\alpha}\|^2|\cdot\right\}\\&+\frac{l\delta_{N_r+1}\underline{\lambda}(R)}{3(l+1)}\mathrm{E}\Big\{\big\|\hat{e}_{t-N+\alpha}\big\|^2\big|\cdot\Big\}-\frac{l\rho_\varepsilon}{l+1}\\&-\varphi_{N_r+1}\bar{\lambda}(R)\frac{N_r\left(\bar{\lambda}(Q)+2\varphi_{N_r+1}\bar{\lambda}(R)\right)\mu_w}{l+1}\\&-\frac{lN_r\varphi_{N_r+1}}{l+1}\bar{\lambda}(R)\mu_w-\frac{l(N_r+1)}{l+1}\bar{\lambda}(R)\mu_v.\end{aligned}\tag{32}$$

By combining (26) and (32), the following inequality holds:

$$\begin{aligned}&2\bar{\lambda}(P)\mathrm{E}\left\{\|\bar{e}_{t-N+\alpha}\|^2|\cdot\right\}\\&\ge\left(\frac{1}{2}\underline{\lambda}(P)+\frac{l\delta_{N_r+1}\underline{\lambda}(R)}{3(l+1)}\right)\mathrm{E}\Big\{\big\|\hat{e}_{t-N+\alpha}\big\|^2\big|\cdot\Big\}-\frac{l\rho_\varepsilon}{l+1}\\&\quad-\varphi_{N_r+1}\bar{\lambda}(R)\frac{N_r\left(\bar{\lambda}(Q)+2\varphi_{N_r+1}\bar{\lambda}(R)\right)\mu_w}{l+1}\\&\quad-\frac{lN_r\varphi_{N_r+1}}{l+1}\bar{\lambda}(R)\mu_w-\frac{l(N_r+1)}{l+1}\bar{\lambda}(R)\mu_v-N_r\bar{\lambda}(Q)\mu_w\\&\quad-(N_r+1)\bar{\lambda}(R)\mu_v.\end{aligned}\tag{33}$$

To simplify (33), let

$$\begin{aligned}f&=\frac{1}{2}\underline{\lambda}(P)+\frac{l\delta_{N_r+1}\underline{\lambda}(R)}{3(l+1)},g=2\bar{\lambda}(P),\\d&=\frac{l\rho_\varepsilon}{l+1}+\varphi_{N_r+1}\bar{\lambda}(R)\frac{N_r\left(\bar{\lambda}(Q)+2\varphi_{N_r+1}\bar{\lambda}(R)\right)\mu_w}{l+1}\\&\quad+\frac{lN_r\varphi_{N_r+1}}{l+1}\bar{\lambda}(R)\mu_w+\frac{l(N_r+1)}{l+1}\bar{\lambda}(R)\mu_v\\&\quad+N_r\bar{\lambda}(Q)\mu_w+(N_r+1)\bar{\lambda}(R)\mu_v.\end{aligned}$$

Then

$$\mathrm{E}\left\{\|\hat{e}_{t-N+\alpha}\|^2|\cdot\right\}\le\frac{g}{f}\mathrm{E}\left\{\|\bar{e}_{t-N+\alpha}\|^2|\cdot\right\}+\frac{d}{f}.\tag{34}$$

Recalling that $\bar{e}_{t-N+\alpha}=x_{t-N+\alpha}-\bar{x}_{t-N+\alpha}$, we have

$$\begin{aligned}&\mathrm{E}\left\{\|\bar{e}_{t-N+\alpha}\|^2|\cdot\right\}\\&=\mathrm{E}\Big\{\big\|A_{r_{t-N+\alpha-1}}x_{t-N+\alpha-1}+w_{t-N+\alpha-1}\\&\quad-A_{\hat{r}_{t-N+\alpha-1}}\hat{x}_{t-N+\alpha-1|t-1}-\hat{w}_{t-N+\alpha-1|t-1}\big\|^2\big|\cdot\Big\}\\&\le3\mathrm{E}\Big\{\big\|\big[A_{r_{t-N+\alpha-1}}-A_{\hat{r}_{t-N+\alpha-1}}\big]x_{t-N+\alpha-1}\big\|^2\big|\cdot\Big\}\\&\quad+\mathrm{E}\{\|w_{t-N+\alpha-1}\|^2\}\\&\quad+3\mathrm{E}\Big\{\big\|A_{\hat{r}_{t-N+\alpha-1}}(x_{t-N+\alpha-1}-\hat{x}_{t-N+\alpha-1|t-1})\big\|^2\big|\cdot\Big\}\\&\quad+3\mathrm{E}\{\|\hat{w}_{t-N+\alpha-1|t-1}\|^2\}.\end{aligned}\tag{35}$$

As for the term $\mathrm{E}\{\|\hat{w}_{t-N+\alpha-1|t-1}\|^2\}$, we have

$$\begin{aligned}&\mathrm{E}\{\|\hat{x}_{t-N+\alpha-1|t-1}-\bar{x}_{t-N+\alpha-1}\|_P|\cdot\}+\mathrm{E}\{\|\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N-1,t-1}\|^2_{\tilde{\mathcal{Q}}}\}\\&\quad+\mathrm{E}\Big\{\Big\|\mathbf{y}^{\alpha,\beta}_{t-N-1,t-1}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N-1,t})\hat{x}_{t-N+\alpha-1|t-1}\\&\quad-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N-1,t-1})\hat{\mathbf{w}}^{\alpha,\beta+1}_{t-N-1,t-1}\Big\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\\&\le\mathrm{E}\Big\{\big\|\hat{x}_{t-N+\alpha-1}-\bar{x}_{t-N+\alpha-1|t-1}\big\|^2_P\Big\}\\&\quad+\mathrm{E}\{\|\tilde{\mathbf{w}}^{\alpha,\beta+1}_{t-N-1,t-1}\|^2_{\tilde{\mathcal{Q}}}\}\\&\quad+\mathrm{E}\Big\{\Big\|\mathbf{y}^{\alpha,\beta}_{t-N-1,t-1}-\mathbf{H}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N-1,t-1})\hat{x}_{t-N+\alpha-1|t-1}\\&\quad-\mathbf{G}(\hat{\mathbf{r}}^{\alpha,\beta}_{t-N-1,t-1})\tilde{\mathbf{w}}^{\alpha,\beta+1}_{t-N-1,t-1}\Big\|^2_{\tilde{\mathcal{R}}}\Big|\cdot\Big\}\end{aligned}$$

where $(\tilde{\mathbf{w}}^{\alpha,\beta+1}_{t-N-1,t-1})'=\left[w_{t-N+\alpha-1}{}'\vdots\left\{\hat{\mathbf{w}}^{\alpha+1,\beta+1}_{t-N-1,t-1}\right\}'\right]'$. Then, the following inequality holds:

$$\begin{aligned}&\mathrm{E}\{\|\hat{w}_{t-N+\alpha-1|t-1}\|_Q\}\\&\le\mathrm{E}\Big\{\|w_{t-N+\alpha-1}\|^2_Q\Big\}+\mathrm{E}\Big\{\big\|\mathcal{H}(r_{t-N+\alpha-1})x_{t-N+\alpha-1}\\&\quad-\mathcal{H}(\hat{r}_{t-N+\alpha-1|t-1})\hat{x}_{t-N+\alpha-1|t-1}\\&\quad+\mathcal{G}_{t-N+\alpha-1}(\mathbf{r}_{t-N-1,t-1})w_{t-N+\alpha-1})\end{aligned}$$

$$
\begin{aligned}
&- \mathcal{G}_{t-N+\alpha-1}(\hat{\mathbf{r}}_{t-N-1,t-1})\hat{w}_{t-N+\alpha-1|t-1})\big\|_R^2\Big\} \\
&+ 2\mathrm{E}\Big\{\big\|\mathcal{G}_{t-N+\alpha-1}(\mathbf{r}^{0,N}_{t-N-1,t-1})w_{t-N+\alpha-1}\big\|_R^2\Big\} \\
&+ 2\mathrm{E}\Big\{\big\|\mathcal{G}_{t-N+\alpha-1}(\mathbf{r}^{0,N}_{t-N-1,t-1})\hat{w}_{t-N+\alpha-1|t-1}\big\|_R^2\Big\}. \\
\le\ & \mathrm{E}\Big\{\|w_{t-N+\alpha-1}\|_Q^2\Big\} + 4\mathrm{E}\Big\{\big\|(\mathcal{H}(r_{t-N+\alpha-1}) \\
&- \mathcal{H}(\hat{r}_{t-N+\alpha-1|t-1}))x_{t-N+\alpha-1}\big\|_R\Big|\cdot\Big\} \\
&+ 4\mathrm{E}\Big\{\big\|\mathcal{H}(\hat{r}_{t-N+\alpha-1|t-1}))(x_{t-N+\alpha-1} \\
&- \hat{x}_{t-N+\alpha-1|t-1})\big\|_R\Big|\cdot\Big\} \\
&+ 2\mathrm{E}\Big\{\big\|\mathcal{H}(\mathbf{r}_{t-N+\alpha-1})w_{t-N+\alpha-1}\big\|_R^2\Big\} \\
&+ 4\mathrm{E}\Big\{\big\|\mathcal{H}(\hat{r}_{t-N+\alpha})\hat{w}_{t-N+\alpha-1|t-1}\big\|_R^2\Big\} \\
&+ 2\mathrm{E}\Big\{\big\|\mathcal{H}(\hat{r}_{t-N+\alpha})w_{t-N+\alpha-1|t-1}\big\|_R^2\Big\}
\end{aligned}
$$

where $\mathcal{H}(r) := C_r\mathcal{F}(r),\ r \in \mathbb{M}$. Let $b := \max_{r\in\mathbb{M}} \|C_r\|^2$. Then

$$
\begin{aligned}
&\mathrm{E}\big\{\|\hat{w}_{t-N+\alpha-1|t-1}\|^2\big\} \\
&\le \frac{4b\rho_{\chi_1}\overline{\lambda}(R)}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)} + \frac{\overline{\lambda}(R)\mathrm{E}\{\|\hat{e}_{t-N+\alpha-1}\|^2|\cdot\}}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)} \\
&\quad + \frac{\overline{\lambda}(Q)+4bu\overline{\lambda}(R)}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)}\mu_w. \qquad (36)
\end{aligned}
$$

Combining (36) and (35), we have the inequality

$$
\begin{aligned}
&\mathrm{E}\left\{\|\bar{e}_{t-N+\alpha}\|^2|\cdot\right\} \\
&\le 3\left(\rho_{\chi_1} + \frac{4b\rho_{\chi_1}\overline{\lambda}(R)}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)}\right) \\
&\quad + 3\left(u + \frac{\overline{\lambda}(R)}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)}\right)\mathrm{E}\{\|\hat{e}_{t-N+\alpha-1}\|^2|\cdot\} \\
&\quad + 3\left(\frac{\overline{\lambda}(Q)+4bu\overline{\lambda}(R)}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)} + \frac{1}{3}\right)\mu_w. \qquad (37)
\end{aligned}
$$

It follows that (34) can be updated as

$$
\mathrm{E}\left\{\|\hat{e}_{t-N+\alpha}\|^2|\cdot\right\} \le \frac{g_0}{f_0}\mathrm{E}\left\{\|\hat{e}_{t-N+\alpha-1}\|^2|\cdot\right\} + \frac{d_0}{f_0}
$$

with

$$
\begin{aligned}
f_0 &= f, \quad g_0 = 3\left(u + \frac{\overline{\lambda}(R)}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)}\right)g, \\
d_0 &= 3\Bigg(\rho_{\chi_1} + \frac{4gb\rho_{\chi_1}\overline{\lambda}(R)}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)} + \Bigg(\frac{\overline{\lambda}(Q)+4bu\overline{\lambda}(R)}{\underline{\lambda}(Q)-4bu\overline{\lambda}(R)} \\
&\quad + \frac{1}{3}\Bigg)\mu_w\Bigg) + d.
\end{aligned}
$$

If P, Q, and R are chosen such that $\frac{g_0}{f_0} < 1$, then the upper bound of the expected value of $\mathrm{E}\big\{\|\hat{e}_{t-N+\alpha}\|^2|x_{t-N}, \mathbf{r}_{t-N,t}\big\}$ can be calculated from

$$
\xi_t = \frac{g_0}{f_0}\xi_{t-1} + \frac{d_0}{f_0},\ \ \xi_\alpha = \bar{d}_0
$$

where $\bar{d}_0$ is defined as the upper bound of $\mathrm{E}\left\{\|\hat{e}_\alpha\|^2|\cdot\right\}$. Then, with

$$
\begin{aligned}
\mathrm{E}\left\{\|\hat{e}_\alpha\|^2|\cdot\right\} &\le \frac{g}{f}\mathrm{E}\left\{\|\bar{e}_\alpha\|^2|\cdot\right\} + \frac{d}{f} \\
&\le \frac{g}{f}\mathrm{E}\Big\{\Big\|\mathcal{F}(\mathbf{r}_{0,\alpha-1})x_0 - \mathcal{F}(\hat{\mathbf{r}}_{0,\alpha-1})\bar{x}_0 \\
&\quad + \mathcal{B}(r_{0,\alpha-1})\mathbf{w}_{0,\alpha-1}\Big\|^2\Big|\cdot\Big\} + \frac{d}{f} \\
&\le \frac{g}{f}\left\{2\rho_{\chi_\alpha} + 2c_\alpha\overline{\lambda}(P_0) + \alpha\eta_\alpha\mu_w\right\} + \frac{d}{f}.
\end{aligned}
$$

The upper bound of $\mathrm{E}\left\{\|\hat{e}_\alpha\|^2|\cdot\right\}$ reduces to

$$
\bar{d}_0 = \frac{g}{f}\left\{2\rho_{\chi_\alpha} + 2c_\alpha\overline{\lambda}(P_0) + \alpha\eta_\alpha\mu_w\right\} + \frac{d}{f}.
$$

E. Proof of Theorem 5.2

Proof: The result of Theorem 5.2 can be obtained using the similar techniques for proving Theorem 5.1 by finding the upper bound and lower bound of objective function $J(\bar{x}_{t-N+\alpha}, \hat{\mathbf{r}}^t_{t-N+\alpha}, \mathbf{y}^t_{t-N+\alpha}, \zeta)$, we can derive the stability results for the EMD-MHE algorithm. ■

F. Proof of Theorem 5.3

Proof: The desired result can be derived similarly to those for Theorems 5.1 and 5.2. From the objective function of the IMD-MHE algorithm, we can find its upper bound and lower bound, leading to the stability results for the IMD-MHE estimator given in Theorem 5.3.

References

[1] J. Xu, P. Shi, C. C. Lim, C. Cai, and Y. Zou, "Reliable tracking control for under-actuated quadrotors with wind disturbances," *IEEE Trans. Syst., Man, Cybern.: Syst.*, doi: 10.1109/TSMC.2017.2782662.

[2] J. Choi and C. C. Lim, "A Cholesky factorization based approach for blind FIR channel identification," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1730–1735, Apr. 2008.

[3] C. Peng and J. Zhang, "Delay-distribution-dependent load frequency control of power systems with probabilistic interval delays," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3309–3317, Jul. 2016.

[4] D. Du, B. Qi, M. Fei, and Z. Wang, "Quantized control of distributed event-triggered networked control systems with hybrid wired-wireless networks communication constraints," *Inf. Sci.*, vol. 380, pp. 74–91, 2017.

[5] D. Du, B. Qi, M. Fei, and C. Peng, "Multiple event-triggered h_2/h_∞ filtering for hybrid networked systems with random network-induced delays," *Inf. Sci.*, vol. 325, pp. 393–408, 2015.

[6] L. Wu, X. Su, and P. Shi, "Sliding mode control with bounded $\mathscr{L}_2$ gain performance of Markovian jump singular time-delay systems," *Automatica*, vol. 48, no. 8, pp. 1929–1933, 2012.

[7] L. Zhang, E. K. Boukas, and J. Lam, "Analysis and synthesis of Markov jump linear systems with time-varying delays and partially known transition probabilities," *IEEE Trans. Automat. Control*, vol. 53, no. 10, pp. 2458–2464, Nov. 2008.

[8] P. Shi, Y. Xia, G. P. Liu, and D. Rees, "On designing of sliding-mode control for stochastic jump systems," *IEEE Trans. Automat. Control*, vol. 51, no. 1, pp. 97–103, Jan. 2006.
[9] F. Li, P. Shi, L. Wu, M. V. Basin, and C. C. Lim, "Quantized control design for cognitive radio networks modeled as nonlinear semi-Markovian jump systems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2330–2340, Apr. 2015.
[10] J. Cui, T. Liu, and Y. Wang, "New stability criteria for a class of Markovian jumping genetic regulatory networks with time-varying delays," *Int. J. Innov. Comput. Inf. Control*, vol. 13, no. 3, pp. 809–822, 2017.
[11] Y. Song, J. Yang, T. Yang, and M. Fei, "Almost sure stability of switching Markov jump linear systems," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2639–2643, 2016.
[12] S. Zhao, Y. S. Shmaliy, and F. Liu, "Fast Kalman-like optimal unbiased FIR filtering with applications," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2284–2294, May 2016.
[13] S. Zhao, B. Huang, and F. Liu, "Linear optimal unbiased filter for time-variant systems without apriori information on initial conditions," *IEEE Trans. Automat. Control*, vol. 62, no. 2, pp. 882–887, Feb. 2017.
[14] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 1, pp. 103–123, Jan. 1998.
[15] W. Liu and I. Hwang, "On hybrid state estimation for stochastic hybrid systems," *IEEE Trans. Automat. Control*, vol. 59, no. 10, pp. 2615–2628, Oct. 2014.
[16] C. E. Seah and I. Hwang, "Stochastic linear hybrid systems: Modeling, estimation, and application in air traffic control," *IEEE Trans. Control Syst. Technol.*, vol. 17, no. 3, pp. 563–575, May 2009.
[17] P. Shi, X. Su, and F. Li, "Dissipativity-based filtering for fuzzy switched systems with stochastic perturbation," *IEEE Trans. Automat. Control*, vol. 61, no. 6, pp. 1694–1699, Jun. 2016.
[18] M. J. Olanrewaju, B. Huang, and A. Afacan, "Online composition estimation and experiment validation of distillation processes with switching dynamics," *Chem. Eng. Sci.*, vol. 65, no. 5, pp. 1597–1608, 2010.
[19] M. J. Olanrewaju, B. Huang, and A. Afacan, "Development of a simultaneous continuum and noncontinuum state estimator with application on a distillation process," *AIChE J.*, vol. 58, no. 2, pp. 480–492, 2012.
[20] A. Alessandri, M. Baglietto, and G. Battistelli, "Receding-horizon estimation for switching discrete-time linear systems," *IEEE Trans. Automat. Control*, vol. 50, no. 11, pp. 1736–1748, Nov. 2005.
[21] G. Ferrari-Trecate, D. Mignone, and M. Morari, "Moving horizon estimation for hybrid systems," *IEEE Trans. Automat. Control*, vol. 47, no. 10, pp. 1663–1676, Oct. 2002.
[22] Y. Guo and B. Huang, "Moving horizon estimation for switching nonlinear systems," *Automatica*, vol. 49, no. 11, pp. 3270–3281, 2013.
[23] Q. Sun, C. C. Lim, P. Shi, and F. Liu, "Moving horizon estimation for Markov jump systems," *Inf. Sci.*, vol. 367–368, pp. 143–158, 2016.
[24] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.
[25] C. E. García, D. M. Prett, and M. Morari, "Model predictive control: Theory and practice–a survey," *Automatica*, vol. 25, no. 3, pp. 335–348, 1989.
[26] C. V. Rao, J. B. Rawlings, and D. Q. Maynei, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Trans. Automat. Control*, vol. 48, no. 2, pp. 246–258, Feb. 2003.
[27] A. Alessandri, M. Baglietto, and G. Battistelli, "A maximum-likelihood Kalman filter for switching discrete-time linear systems," *Automatica*, vol. 46, no. 11, pp. 1870–1876, 2010.
[28] A. Logothetis and V. Krishnamurthy, "Expectation maximization algorithms for MAP estimation of jump Markov linear systems," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2139–2156, Aug. 1999.
[29] A. Doucet, A. Logothetis, and V. Krishnamurthy, "Stochastic sampling algorithms for state estimation of jump Markov linear systems," *IEEE Trans. Automat. Control*, vol. 45, no. 2, pp. 188–202, Feb. 2000.
[30] R. Vidal, A. Chiuso, and S. Soatto, "Observability and identifiability of jump linear systems," in *Proc. 41st IEEE Conf. Decis. Control*, Las Vegas, NV, USA, 2002, pp. 3614–3619.
[31] A. Alessandri, M. Baglietto, and G. Battistelli, "Luenberger observers for switching discrete-time linear systems," *Int. J. Control*, vol. 80, no. 12, pp. 1931–1943, 2007.

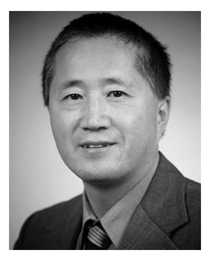
Qing Sun received the B.Sc. degree in automation and the Ph.D. degree in control science and engineering from Jiangnan University, Wuxi, China, in 2011 and 2017, respectively.

From 2014 to 2016, she was a Visiting Student with the School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA, Australia. She is currently with the Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronical Engineering and Automation, Shanghai University, Shanghai, China. Her research interests include state estimation theory, advanced control theory, stochastic signal processing, etc.

Cheng-Chew Lim (SM'02) received the B.Sc. (Hons.) degree in electronic and electrical engineering and the Ph.D. degree in electronic and electrical engineering from Loughborough University, Loughborough, U.K., in 1977 and 1981, respectively.

He is currently a Professor with the University of Adelaide, Adelaide, SA, Australia. His research interests include control and systems theory, autonomous systems, machine learning, and optimization techniques and applications.

Dr. Lim is an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS and an Editorial Board Member for the *Journal of Industrial and Management Optimization*, and was a Guest Editor for a number of journals, including *Discrete and Continuous Dynamical System-Series B*.



Peng Shi (M'95–SM'98–F'15) received the Ph.D. degree in electrical engineering from the University of Newcastle, Callaghan, NSW, Australia, in 1994, and the Ph.D. degree in mathematics from the University of South Australia, Adelaide, SA, Australia, in 1998. He received the Doctor of Science degree from the University of Glamorgan, Pontypridd, U.K., in 2006, and the Doctor of Engineering degree from the University of Adelaide, Adelaide, SA, Australia, in 2015.

He is currently a Professor with the University of Adelaide. His research interests include system and control theory, computational intelligence, and operational research.

Prof. Shi was the recipient of the Andrew Sage Best Transactions Paper Award from the IEEE SMC Society in 2016, and was a Runner-Up of the SCOPUS Researcher of the Year in 2017 (Australia and New Zealand). He was on the editorial board for a number of journals, including *Automatica*, IEEE TRANSACTIONS ON AUTOMATIC CONTROL; IEEE TRANSACTIONS ON CYBERNETICS; IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS; IEEE TRANSACTIONS ON FUZZY SYSTEMS; and IEEE CONTROL SYSTEMS LETTERS. He is currently a Board of Governors Member, IEEE SMC Society, and was an IEEE Distinguished Lecturer; and the Chair of the Control Aerospace and Electronic Systems Chapter, IEEE South Australia Section. He is a Fellow of the Institution of Engineering and Technology, and the Institute of Engineering Australia.

Fei Liu (M'02) received the B.Sc. degree in electrical technology and the M.Sc. degree in industrial automation from the Wuxi Institute of Light Industry, Wuxi, China, in 1987 and 1990, respectively, and the Ph.D. degree in control science and control engineering from Zhejiang University, Hangzhou, China, in 2002.

From 1990 to 1999, he was an Assistant, a Lecturer, and an Associate Professor with the Wuxi Institute of Light Industry. Since 2003, he has been a Professor with the Institute of Automation, Jiangnan University, Wuxi, China. From 2005 to 2006, he was a Visiting Professor with the University of Manchester, Manchester, U.K. His research interests include advanced control theory and applications, batch process control engineering, statistical monitoring and diagnosis in industrial processes, and intelligent technique with emphasis on fuzzy and neural systems.