State estimation over non-acknowledgment networks with Markovian packet dropouts ⋆

Hong Lin [a,b] James Lam [b] Zidong Wang [c] Hak-Keung Lam [d]

[a] *Institute of Intelligence Science and Engineering, Shenzhen Polytechnic, Shenzhen 518055, China.*

[b] *Department of Mechanical Engineering, The University of Hong Kong, Pokfulam, Hong Kong.*

[c] *Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom.*

[d] *Department of Informatics, King's College London, Bush House, Strand Campus, WC2B 4BG, United Kingdom.*

Abstract

In this paper, we investigate state estimation for systems with packet dropouts. According to whether there are acknowledgment (ACK) signals sent by the actuator to the estimator indicating the status of control packet dropouts or not, the systems are classified into two types: ACK systems, those with ACK signals, and non-ACK (NACK) systems, those without. We first obtain the optimal estimator (OE) for NACK systems with Markovian packet dropouts. However, the number of the components in the OE grows exponentially, making its stability analysis complicated and its computation time-consuming. Therefore, we proceed to design a computationally efficient approximate optimal estimator (AOE) using a relative-entropy-based approach. We prove that the proposed AOE has the same stability as the OE. We show that, even the separation principle does not hold for NACK systems, the stability of the OE can also be investigated separately; and discover that the OE for an NACK system has the same stability as the OE for the corresponding ACK system, even their structures are quite different. Finally, for strongly observable NACK systems, we establish a necessary and sufficient condition for the stability of the OE and the AOE.

Key words: optimal estimation; approximate estimation; packet dropout; packet acknowledgment; relative entropy;

1 Introduction

1.1 Research Background and Literature Review

There is no doubt that state estimation is an important topic in both theoretical research and practical applications. In the past decade, a substantial body of literature has been devoted to state estimation for systems with packet dropouts [5, 34]. This is owing to the development of communication and network techniques, which allows the components in control systems being connected over a distant range to share information. Due to network-induced constraints [18, 32, 37] or malicious network attacks [31, 36], data transmitted over networks may be randomly dropped.

Three commonly used models for control systems with packet dropouts are shown in Fig. 1. The key difference between these two model lies in the acknowledgment (ACK) channel. This channel is particularly used by the actuator to send an ACK signal (0 or 1) to inform the estimator whether control packets are dropped or not. A system has an ACK channel in Fig. 1 (b) is called an ACK system, and a system in Fig. 1 (a) without such an ACK channel is called a non-ACK (NACK) system. It is well known [25] that the separation principle—that is, the controller and the estimator can be designed separately—holds for ACK systems but fails for NACK ones [25]. That is why the majority of estimation-related issues for ACK systems can be investigated by using the model in Fig. 1 (c) without considering control inputs.

In real-world applications, a typical example of ACK (resp. NACK) systems is networked systems with a transmission control protocol (TCP) (resp. a user datagram

⋆ The work was supported in part by GRF HKU (17200918), in part by National Natural Science Foundation of China under Grant 61873148 and Grant 61873338, and in part by National Key R&D Program of China under Grant 2018YF-B1700100.
Corresponding author: James Lam.
Email addresses: linhongzju@163.com (Hong Lin), james.lam@hku.hk (James Lam), Zidong.wang@brunel.ac.uk (Zidong Wang), hak-keung.lam@kcl.ac.uk (Hak-Keung Lam).

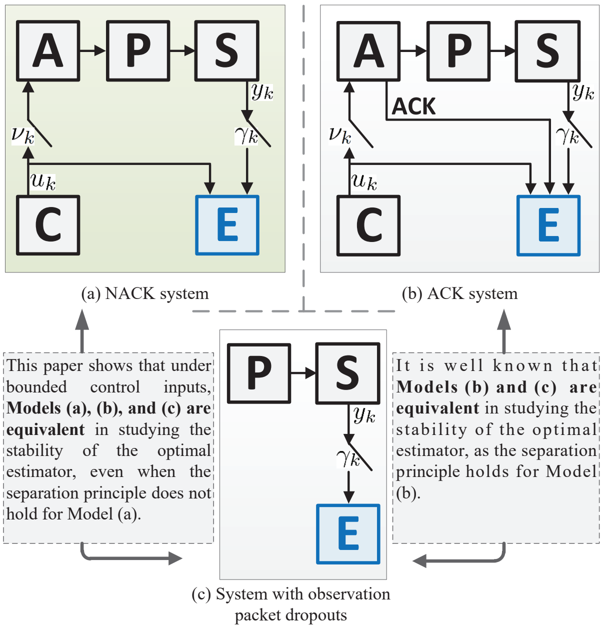

Fig. 1. NACK system, ACK system, and system with only observation packet dropouts. The blocks P, S, E, C, and A denote the plant, sensor, estimator, controller and actuator, respectively.

protocol (UDP)) [15, 25], where TCP has an ACK mechanism for packet dropouts while UDP has not [6]. Under TCP, the sender retransmits the dropped packets until acknowledged by the receiver that the packet has been successfully received. Such an ACK-based mechanism guarantees the success of data delivery, but it may cause transmission delay, since due to continuous packet dropouts, a packet may be retransmitted several times before successfully received. Such ACK systems are suitable for applications that require guaranteed delivery (e.g., industrial manufacturing [19] and process monitoring [2]), where delay is not of the prime concern. UDP, at the cost of some delivery reliability, allows more timely and consistent communication without waiting for ACK signals and then retransmitting the dropped packets, and therefore such NACK systems are preferable choices for real-time networked systems [23].

This paper studies state estimation problems for NACK systems, and the recent advancement of state estimation for ACK and NACK systems is reviewed as follows. For convenience of formulation, we denote an ACK system, an NACK system, and a system in Fig. 1 (c) by $\mathcal{S}^{*}_{\mathsf{ACK}}$, $\mathcal{S}^{*}_{\mathsf{NACK}}$, and $\mathcal{S}^{*}_{\mathsf{Obs}}$, respectively, where the superscript $* =$ $\mathsf{i.i.d}$ (or $* = \mathsf{Markov}$) means the packet dropout follows an i.i.d. Bernoulli distribution (or a Markov process).

Works on ACK systems: The optimal estimator (OE) for an $\mathcal{S}^{\mathsf{i.i.d}}_{\mathsf{Obs}}$ or $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{Obs}}$ system is a time-varying Kalman filter [26, 35], and the OE for $\mathcal{S}^{\mathsf{i.i.d}}_{\mathsf{Obs}}$ systems is stable if the observation packet loss rate is less than a critical value [26]. For $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{Obs}}$ systems, the stability conditions of the OE have been obtained for scalar, second-order, and certain classes of high-order systems in [9, 35], and for diagonalizable non-degenerate systems in [21]. As the separation principle holds for ACK systems, these estimation-related results obtained for $\mathcal{S}^{\mathsf{i.i.d}}_{\mathsf{Obs}}$ and $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{Obs}}$ systems also hold for $\mathcal{S}^{\mathsf{i.i.d}}_{\mathsf{ACK}}$ and $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{ACK}}$ systems, respectively. However, the OE for NACK systems [13, 16] differs a lot from the OE for ACK systems, and the proposed methods in the aforementioned literature for ACK systems are not applicable to NACK systems.

Works on NACK systems: The OE for an $\mathcal{S}^{\mathsf{i.i.d}}_{\mathsf{NACK}}$ system and its stability conditions were first obtained for NACK systems with only control packet dropouts [16], and then extended to systems with both control and observation packet dropouts [13]. The stability of the OE only depends on the observation packet loss rate; however, the OE consists of an exponentially growing number of components, making its computation time-consuming [13]. To address the computational problem, two approximate optimal estimators (AOEs) [3, 12] were developed to compute the optimal estimates for $\mathcal{S}^{\mathsf{i.i.d}}_{\mathsf{NACK}}$ systems with relatively less computational efforts.

1.2 *Underlying issues and research motivation*

To the best of our knowledge, no results have been reported on the OE and the AOE for $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{NACK}}$ systems. Unlike the systems with i.i.d. packet dropouts in which only one factor (the packet dropout rate) affects packet dropouts, two temporally correlated factors (packet recovery and failure rates) govern Markovian packet dropouts. It is pointed out in [35] that for ACK systems, due to the temporal correlation, the methods developed for the i.i.d. packet dropout case are not applicable to the Markovian case, and the estimator stability analysis for the Markovian case is more challenging. Similarly, we found that these two factors and the temporal correlation also bring technical difficulties in applying the methods proposed for $\mathcal{S}^{\mathsf{i.i.d}}_{\mathsf{NACK}}$ systems to $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{NACK}}$ ones.

- For the OE, as far as we know, there are two methods [13, 16] available for analyzing the estimator stability for $\mathcal{S}^{\mathsf{i.i.d}}_{\mathsf{NACK}}$ systems. The temporal correlation makes it difficult to construct well-defined pdfs as in [13, 16] with desired properties for analyzing the OE stability for $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{NACK}}$ systems.
- For the AOE, the existing methods [3, 12] are not applicable to $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{NACK}}$ systems either, since (i) The method proposed in [3] assumes that there is a maximal number for consecutive packet dropouts, but we allow an arbitrary number of consecutive packet dropouts. Thus, the assumption does not hold for the system we considered. (ii) The results in [4, 7, 33] suggest that "when the system parameters which cannot be observed directly (like the status of the control packet dropout in an $\mathcal{S}^{\mathsf{Markov}}_{\mathsf{NACK}}$ system) evolve according to Markov processes, the error covariance

of the LMMSE estimator does not necessarily has a Riccati-equation-like recursive form." Hence, it would be infeasible for $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ systems to develop a recursive linear minimum mean square error (LMMSE) estimator, like the one developed in [12], to study the stability of an AOE.
- The aforementioned methods involve dealing with the probability of the control packet dropout status. This quantity is governed by one factor (the packet dropout rate) in $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{i.i.d}}$ systems, and it—that is, $\overline{\alpha}_k^{[i]}$ in (15)—contains two temporally correlated factors and becomes more complicated in $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ systems, which makes it difficult to extend these methods to $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ systems.

It can be seen from the discussion above that for $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ systems, these two temporally correlated factors make the estimator stability analysis more difficult than that for $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{i.i.d}}$ systems. To deal with the temporal correlation in Markov processes, continuous efforts [9, 21, 24, 35] have been devoted to developing new techniques for ACK systems in the last decade, but few method is proposed for NACK systems. It motivates us to explore new methods for $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ systems to address the following two problems: **Problem 1:** What are the stability conditions for the OE? How do the packet recovery/failure rates of observation and control inputs affects the OE stability? **Problem 2:** How to design an AOE with a good estimation performance?

1.3 Main results and contributions

To our best knowledge, this paper is the first attempt to investigate the estimation issues for NACK systems with Markovian packet dropouts. Our main results and contributions are summarized as follows:

1) From the OE perspective, (i) we obtain the OE for an $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ system, and show that even when the separation principle does not hold for NACK systems and the OE for $\mathcal{S}_{\mathsf{Obs}}^{\mathsf{Markov}}$ systems differs a lot from the OE for $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ systems, interestingly, the estimator stability for $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ systems can also be analyzed by using the model $\mathcal{S}_{\mathsf{Obs}}^{\mathsf{Markov}}$ in Fig. 1 (c). (ii) We show that for a strongly observable $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ system, the OE stability is independent of the observation packet failure rate and the control packet recovery/failure rate; and the OE is stable if and only if the observation packet recovery rate is greater than a threshold value.
2) From an approximation estimation perspective, we develop a relative-entropy-based (RE-based) AOE for $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ systems. It needs to mention that relative entropy (also known as Kullback-Leibler divergence) has been utilized to reduce the components of a Gaussian mixture in various fields, from stationary data sets [20, 30] to dynamic systems [22, 29]. However, usually, no analytical solutions are available to RE-based Gaussian mixture reduction problems [8], which has to be solved numerically. Consequently, the stability/performance of RE-based methods is usually evaluated by simulation methods and difficult to be theoretically determined. We prove that the proposed RE-based AOE has the same stability as the OE.

The rest of the paper is organized as follows: In Section 2, the system setup is introduced. The OE and an AOE are proposed in Sections 3 and 4, respectively. The performance and stability of the OE/AOE are studied in Sections 5 and 6, respectively. In Section 7, numerical examples are given to illustrate our main results. The conclusions are presented in Section 8. The proofs of the lemmas are given in the appendix section.

Notation:

- $p(\cdot)$ denotes a probability density function (pdf).
- $\mathcal{N}(\mu, P)$ denotes a Gaussian pdf of with the mean μ and the covariance P.
- $\mathbb{P}(\cdot)$, $\mathbb{E}[\cdot]$, $\mathbb{C}\mathrm{ov}[\cdot]$ stand for a probability measure, mathematical expectation, and covariance, respectively. $\mathbb{E}_y[\cdot]$ means that the mathematical expectation is taken with respect to the random variable y.
- $(\cdot)'$ denotes the transpose of a matrix or a vector.
- $(\cdot)_I^2$ with the identity matrix I denotes $(\cdot)(\cdot)'$.
- $\mathrm{tr}(M)$ denotes the trace of a matrix M.
- $(\cdot)_2$ denotes the binary representation, e.g., $(11)_2 = 3$.
- $M \in \mathbb{M}_+^n$ or $M > 0$ denotes M is a positive definite matrix.
- $\mathbb{i}_a^b$ denotes the set $\{i \in \mathbb{N} | a \leq i \leq b\}$, where $\mathbb{N}$ is the set of natural numbers.
- For a matrix sequence M_k, $k \in \mathbb{N}$, $\sup M_k$ and $\inf M_k$ denote the supremum and the infimum of the sequence M_k, respectively.

2 System Setup and Preliminaries

2.1 System setup

Consider the following discrete-time linear system

$$\begin{aligned} x_k &= Ax_{k-1} + \nu_k Bu_k + \omega_k \\ y_k &= Cx_k + \upsilon_k \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the system state, $u_k \in \mathbb{R}^{n_u}$ is the control input, $y_k \in \mathbb{R}^{n_y}$ is the observation. ω_k and υ_k are zero mean Gaussian noises with covariances $Q > 0$ and $R > 0$, respectively.

NACK system: The NACK system considered in this paper is the one in (1) with the system model in Fig. 1 (a). For an NACK system, say $\mathcal{S}$, the *corresponding ACK system* refers to the one, with the system model in Fig. 1 (b), has the same parameters as the NACK system $\mathcal{S}$.

Communication channels: The packet dropouts over the controller-to-actuator (C-A) and the sensor-to-estimator (S-E) channels are modeled by Markov processes $\{\nu_k\}$ and $\{\gamma_k\}$, respectively. Define $\Gamma_k \triangleq \{\gamma_1, \ldots, \gamma_k\}$ and $\Upsilon_k \triangleq \{\nu_1, \ldots, \nu_k\}$.

Their transition matrices are the following:

$$\begin{bmatrix} \mathbb{P}(\gamma_{k+1}=0|\gamma_k=0) & \mathbb{P}(\gamma_{k+1}=1|\gamma_k=0) \\ \mathbb{P}(\gamma_{k+1}=0|\gamma_k=1) & \mathbb{P}(\gamma_{k+1}=1|\gamma_k=1) \end{bmatrix} = \mathcal{P}$$

$$\begin{bmatrix} \mathbb{P}(\nu_{k+1}=0|\nu_k=0) & \mathbb{P}(\nu_{k+1}=1|\nu_k=0) \\ \mathbb{P}(\nu_{k+1}=0|\nu_k=1) & \mathbb{P}(\nu_{k+1}=1|\nu_k=1) \end{bmatrix} = \mathcal{Q}$$

where $\mathcal{P} = \begin{bmatrix} 1-p_1 & p_1 \\ p_2 & 1-p_2 \end{bmatrix}$ and $\mathcal{Q} = \begin{bmatrix} 1-q_1 & q_1 \\ q_2 & 1-q_2 \end{bmatrix}$. p_1 and q_1 are called packet recovery rates; p_2 and q_2 are called packet failure rates.

Estimator: All the observations that actually received by the estimator up to time k is $Y_k \triangleq \{y_j | \gamma_j = 1, 1 \le j \le k\}$. This paper will, based on the received observations Y_k, derive the OE and then design an AOE.

Controller: This paper, like the works [13, 26], only focuses on the estimation problems and is not intended to design controllers and estimators in parallel. For the controller, we assume, like the case in [13, 16], that control inputs u_k are deterministic and bounded. Denote the upper bound of $Bu_k u_k' B'$ by U.

Assumption 1 *$(A, Q^{1/2})$ is controllable and (A, C) is observable. ω_k, υ_k, ν_k, γ_k, and $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$ are mutually independent. The Markov chains ν_k and γ_k are irreducible and stationary.*

2.2 Preliminaries

Lemma 1 *[28, p. 88, (3.7)–(3.14); p. 98] Let $X = AZ + Bu + W$ and $Y = CX + V$, where $Z \sim \mathcal{N}(\overline{Z}, P_Z)$, $W \sim \mathcal{N}(0, Q)$, $V \sim \mathcal{N}(0, R)$, and A, B, C, u are constants. Then*

$$p(X) = \mathcal{N}(\overline{X}, \overline{P}_X) \tag{2a}$$

$$p(X|Y) = \mathcal{N}(\overline{X} + K(Y - C\overline{X}), (I - KC)\overline{P}_X) \tag{2b}$$

$$p(Y) = \mathcal{N}(C\overline{X}, C\overline{P}_X C' + R) \tag{2c}$$

where $\overline{X} = A\overline{Z} + Bu, \overline{P}_X = AP_Z A' + Q, K = \overline{P}_X C'(C\overline{P}_X C' + R)^{-1}$.

Define some operators as follows:

$$\overline{g}(P, \gamma, X, Y) = APA' + X - \gamma APC'(CPC' + Y)^{-1}CPA'$$

$$g(P, \gamma) = \overline{g}(P, \gamma, Q, R)$$

$$h(P, K) = (I - KC)P(I - KC)' + KRK'$$

$$\phi(P, \gamma) = P - \gamma PC'(CPC' + R)^{-1}CP \tag{3}$$

Lemma 2 *The following facts hold.*

(i) $\forall K, h(P, K_P) \le h(P, K)$*, where* $K_P = PC'(CPC' + R)^{-1}$.

(ii) $\phi(P, 1) = h(P, K_P)$.

(iii) *If* $X \le Y$*, then* $g(X, r) \le g(Y, r)$, $h(X, K) \le h(Y, K)$*, and* $\phi(X, \gamma) \le \phi(Y, \gamma)$, $\gamma = 0$ *or* 1.

(iv) *If* $X \le Y, Q_1 \le Q_2, R_1 \le R_2$*, then* $\overline{g}(X, \gamma, Q, R) \le \overline{g}(Y, \gamma, Q, R)$, $\overline{g}(X, \gamma, Q_1, R) \le \overline{g}(X, \gamma, Q_2, R)$*, and* $\overline{g}(X, \gamma, Q, R_1) \le \overline{g}(X, \gamma, Q, R_2)$.

The OE for an $\mathcal{S}_{\mathsf{ACK}}^{\mathsf{Markov}}$ system has been obtained in [21]. Denote its prediction and estimation error covariances by $\overline{P}_k^{\mathsf{ACK}}$ and P_k^{ACK}, respectively. From [21, Eqs. (2)–(5)], it follows that

$$\overline{P}_k^{\mathsf{ACK}} = AP_{k-1}^{\mathsf{ACK}} A' + Q \tag{4a}$$

$$P_k^{\mathsf{ACK}} = \phi(\overline{P}_k^{\mathsf{ACK}}, \gamma_k), \text{ with } P_0^{\mathsf{ACK}} = P_0. \tag{4b}$$

3 Optimal Estimation

In this section, we derive the OE for an $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ system.

Definition 1 (Optimal estimate) *The optimal estimate $\widehat{x}_k$ is the one that minimizes $\mathbb{E}_{x_k}\left[(x_k - \widehat{x}_k)_I^2 | Y_k\right]$.*

It is well known [1] that the desired optimal estimate $\widehat{x}_k = \mathbb{E}[x_k|Y_k]$. Thus, we derive the conditional pdf $p(x_k|Y_k)$ as follows.

3.1 Representation of control packet dropouts

The random variables $\{\nu_k, \ldots, \nu_1\}$, where $\nu_j = 0$ or 1 and $1 \le j \le k$, have 2^k different values, which form a probability space, denoted by Ω_k. Define a one-to-one mapping $\psi : \Omega_k \to \mathbb{N}$ by $j = \psi(\nu_k, \ldots, \nu_1) = (\nu_k \ldots \nu_1)_2 + 1$, that is, $j = \nu_k 2^{k-1} + \ldots + \nu_1 2^0 + 1$. By ψ, we can denote an element $(\nu_k, \ldots, \nu_1) \in \Omega_k$ by $\Upsilon_k^{[i]}$ as follows: For $1 \le i \le 2^k$,

$$\Upsilon_k^{[i]} = (\nu_k, \ldots, \nu_1) \in \Omega_k \text{ with } i = \psi(\nu_k \ldots \nu_1).$$

It is easy to verify that the following equalities hold: for $1 \le i \le 2^n$,

$$\Upsilon_n^{[i]} = \begin{cases} \{\nu_n = 0, \Upsilon_{n-1}^{[i]}\}, & i \in \mathbb{i}_1^{2^{n-1}} \\ \{\nu_n = 1, \Upsilon_{n-1}^{[i-2^{n-1}]}\}, & i \in \mathbb{i}_{2^{n-1}+1}^{2^n}. \end{cases} \tag{5}$$

Lemma 3 *The following facts hold:*

(i) For $j_1, j_2 \in \{0,1\}$,

$$p(\nu_{n+1}=j_2|\nu_n=j_1,\Upsilon_{n-1}^{[i]})=p(\nu_{n+1}=j_2|\nu_n=j_1) \quad (6)$$

$$p(\nu_{n+1}=j_1|\Upsilon_n^{[i]},Y_n)=p(\nu_{n+1}=j_1|\Upsilon_n^{[i]}) \quad (7)$$

(ii)

$$p(\nu_{n+1}=0|\Upsilon_n^{[i]}) = \begin{cases} (1-q_1), & i \in \mathbb{i}_1^{2^{n-1}} \\ q_2, & i \in \mathbb{i}_{2^{n-1}+1}^{2^n} \end{cases}$$

$$p(\nu_{n+1}=1|\Upsilon_n^{[i]}) = \begin{cases} q_1, & i \in \mathbb{i}_1^{2^{n-1}} \\ (1-q_2), & i \in \mathbb{i}_{2^{n-1}+1}^{2^n} \end{cases}$$

3.2 Conditional probability density functions of x_k

By using the total probability law,

$$p(x_k|Y_k) = \sum_{i=1}^{2^k} p(x_k|Y_k,\Upsilon_k^{[i]})p(\Upsilon_k^{[i]}|Y_k) \quad (8)$$

$$p(x_{k+1}|Y_k) = \sum_{i=1}^{2^{k+1}} p(x_{k+1}|Y_k,\Upsilon_{k+1}^{[i]})p(\Upsilon_{k+1}^{[i]}|Y_k) \quad (9)$$

The conditional pdfs on the right-hand side of (8) and (9) are obtained in Lemmas 4 and 5.

Lemma 4

$$p(x_k|Y_k,\Upsilon_k^{[i]}) = \mathcal{N}(\widehat{m}_k^{[i]},\widehat{M}_k) \quad (10)$$

$$p(x_{k+1}|Y_k,\Upsilon_{k+1}^{[i]}) = \mathcal{N}(\overline{m}_{k+1}^{[i]},\overline{M}_{k+1}), \quad (11)$$

where

$$\begin{aligned} \overline{m}_{k+1}^{[i]} &= \begin{cases} A\widehat{m}_k^{[i]}, & i \in \mathbb{i}_1^{2^{n-1}} \\ A\widehat{m}_k^{[i-2^{n-1}]} + Bu_{k+1}, & i \in \mathbb{i}_{2^{n-1}+1}^{2^n} \end{cases} \\ \overline{M}_{k+1} &= A\widehat{M}_kA' + Q \quad (12) \\ K_{k+1} &= \overline{M}_{k+1}C'(C\overline{M}_{k+1}C'+R)^{-1} \\ \widehat{m}_{k+1}^{[i]} &= \overline{m}_{k+1}^{[i]} + \gamma_{k+1}K_{k+1}(y_{k+1}-C\overline{m}_{k+1}^{[i]}) \\ \widehat{M}_{k+1} &= (I-\gamma_{k+1}K_{k+1}C)\overline{M}_{k+1} \quad (13) \end{aligned}$$

with $\widehat{m}_0^{[1]} = \overline{x}_0$ and $\widehat{M}_0 = P_0$.

Lemma 5 *Let $\overline{\alpha}_{n+1}^{[i]} \triangleq p(\Upsilon_{n+1}^{[i]}|Y_n)$ and $\widehat{\alpha}_n^{[i]} \triangleq p(\Upsilon_n^{[i]}|Y_n)$*

$$\widehat{\alpha}_n^{[i]} = \left(\frac{\mu_n^{[i]}}{\sum_{i=1}^{2^n}\mu_n^{[i]}\overline{\alpha}_n^{[i]}}\right)^{\gamma_n} \cdot \overline{\alpha}_n^{[i]} \quad (14)$$

$$\overline{\alpha}_{n+1}^{[i]} = \begin{cases} \widehat{\alpha}_n^{[i]}(1-q_1), & i \in \mathbb{i}_1^{2^{n-1}} \\ \widehat{\alpha}_n^{[i]}q_2, & i \in \mathbb{i}_{2^{n-1}+1}^{2\cdot 2^{n-1}} \\ \widehat{\alpha}_n^{[i-2^n]}q_1, & i \in \mathbb{i}_{2\cdot 2^{n-1}+1}^{3\cdot 2^{n-1}} \\ \widehat{\alpha}_n^{[i-2^n]}(1-q_2), & i \in \mathbb{i}_{3\cdot 2^{n-1}+1}^{4\cdot 2^{n-1}} \end{cases} \quad (15)$$

where $\mu_n^{[i]} \triangleq p(y_n|\Upsilon_n^{[i]},Y_{n-1})$.

3.3 Optimal estimator for NACK systems

From (8), (10), and $\widehat{\alpha}_k^{[i]}$ in Lemma 5, it follows that

$$p(x_k|Y_k) = \sum_{i=1}^{2^k}\widehat{\alpha}_k^{[i]}\mathcal{N}(\widehat{m}_k^{[i]},\widehat{M}_k). \quad (16)$$

Theorem 1 (Optimal estimator) *The OE for an $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ system in (1) is the following:*

$$\widehat{x}_k = \sum_{i=1}^{2^k}\widehat{\alpha}_k^{[i]}\widehat{m}_k^{[i]},\ P_k = \widehat{M}_k + \sum_{i=1}^{2^k}\widehat{\alpha}_k^{[i]}(\widehat{m}_k^{[i]}-\widehat{x}_k)_I^2, \quad (17)$$

where $\widehat{\alpha}_k^{[i]}, \widehat{m}_k^{[i]}$, and $\widehat{M}_k$ can be computed by Lemmas 4 and 5. Moreover, $P_k^{\mathsf{ACK}} = \widehat{M}_k$.

Proof: For a Gaussian mixture like (16), the mean $\widehat{x}_k$ and the covariance P_k take the forms as in (17), which is the existing result in [1, p. 213]. $P_k^{\mathsf{ACK}} = \widehat{M}_k$ can be proved by noting that P_k^{ACK} in (4a)(4b) evolves in the same way as $\widehat{M}_k$ in (12)(13), with $P_0^{\mathsf{ACK}} = \widehat{M}_0 = P_0$.

4 Approximate Optimal Estimator

In this section, we first develop a relative-entropy-based approach to reduce the number of the components in $p(x_k|Y_k)$, and then propose a computationally efficient algorithm to approximately compute the optimal estimate.

4.1 Relative-entropy-based Gaussian mixture reduction

Relative entropy, a well-known measure of the deviation between two pdfs, and the relative-entropy-based Gaussian mixture reduction (RE-based GM reduction) problem are introduced as follows [10].

Definition 2 (Relative entropy) *For two pdfs $f_1(x)$ and $f_2(x)$, $x \in \mathbb{R}^n$, the* **relative entropy** *of f_1 with respect to f_2, denoted $\mathscr{D}(f_1,f_2)$, is defined as*

$$\mathscr{D}(f_1,f_2) \triangleq \int_{\mathbb{R}^n} f_1(x)\log(f_1(x)/f_2(x))\mathrm{d}x.$$

Definition 3 (RE-based GM reduction) *For a given Gaussian mixture $f_1(x) = \sum_{i=1}^{N} \alpha_i \mathcal{N}(\mu_i, P_i)$, $x \in \mathbb{R}^n$, the* **RE-based GM reduction problem** *is to determine a Gaussian mixture $f_2(x) = \sum_{i=1}^{M} \alpha_i^* \mathcal{N}(\mu_i^*, P_i^*)$ with $M < N$, such that $\mathscr{D}(f_1, f_2)$ is minimized. That is, to solve the following optimization problem:*

$$\mathscr{D}\big(f_1, \sum_{i=1}^{M} \alpha_i^* \mathcal{N}(\mu_i^*, P_i^*)\big) \triangleq \min_{\alpha_i^\sharp, \mu_i^\sharp, P_i^\sharp} \mathscr{D}\big(f_1, \sum_{i=1}^{M} \alpha_i^\sharp \mathcal{N}(\mu_i^\sharp, P_i^\sharp)\big)$$

where $\mu_i^\sharp \in \mathbb{R}^n, P_i^\sharp \in \mathbb{M}_+^n, \alpha_i^\sharp \in [0,1]$ with $\sum_{i=1}^{M} \alpha_i^\sharp = 1$.

Lemma 6 *For a given N-component Gaussian mixture $p(x) = \sum_{i=1}^{N} \alpha_i \mathcal{N}(m_i, P_i)$, $x \in \mathbb{R}^n$, the optimization problem*

$$\mathscr{D}\Big(p(x), \mathcal{N}(m^*, P^*)\Big) = \min_{m \in \mathbb{R}^n, P \in \mathbb{M}_+^n} \mathscr{D}\Big(p(x), \mathcal{N}(m, P)\Big)$$

has a unique solution that

$$m^* = \sum_{i=1}^{N} \alpha_i m_i, \;\; P^* = \sum_{i=1}^{N} \alpha_i \big(P_i + (m^* - m_i)_I^2\big). \quad (18)$$

Lemma 7 *Let $\overline{\nu}_k^{[i]} \triangleq p(\nu_k = i | Y_{k-1})$ and $\widehat{\nu}_k^{[i]} \triangleq p(\nu_k = i | Y_k)$, for $i = 0$ or 1. Then,*

$$\overline{\nu}_k^{[0]} = (1 - q_1)\widehat{\nu}_{k-1}^{[0]} + q_2 \widehat{\nu}_{k-1}^{[1]} \quad (19a)$$
$$\overline{\nu}_k^{[1]} = q_1 \widehat{\nu}_{k-1}^{[0]} + (1 - q_2)\widehat{\nu}_{k-1}^{[1]} \quad (19b)$$
$$\widehat{\nu}_k^{[i]} = \left(\frac{1}{c} \varpi_k^{[i]}\right)^{\gamma_k} \overline{\nu}_k^{[i]} \quad (19c)$$

where $\varpi_k^{[i]} \triangleq p(y_k | \nu_k = i, Y_k)$ and $c = \sum_{i=0}^{1} \varpi_k^{[i]} \overline{\nu}_k^{[i]}$.

Lemma 8 *If $p(x_{k-1}|Y_{k-1}) = \mathcal{N}(\widehat{x}_{k-1}^{\mathsf{int}}, P_{k-1}^{\mathsf{int}})$, then $p(x_k | \nu_k = i, Y_k) = \mathcal{N}(\widehat{z}_k^{[i]}, \widehat{Z}_k)$, with $i = 0$ or 1, where $\widehat{z}_k^{[i]}$ and $\widehat{Z}_k$ are computed as follows:*

$$\overline{z}_k^{[i]} = A\widehat{x}_{k-1}^{\mathsf{int}} + iBu_k \quad (20a)$$
$$\overline{Z}_k = AP_{k-1}^{\mathsf{int}}A' + Q \quad (20b)$$
$$\widehat{z}_k^{[i]} = \overline{z}_k^{[i]} + K_k^\varepsilon(y_k - C\overline{z}_k^{[i]}) \quad (20c)$$
$$\widehat{Z}_k = (I - K_k^\varepsilon C)\overline{Z}_k(I - K_k^\varepsilon C)' + K_k^\varepsilon R (K_k^\varepsilon)' \quad (20d)$$

where $K_k^\varepsilon = \gamma_k \overline{Z}_k C'(C\overline{Z}_k C' + R)^{-1}$.

Lemma 9 *If $p(x_{k-1}|Y_{k-1}) = \mathcal{N}(\widehat{x}_{k-1}^{\mathsf{int}}, P_{k-1}^{\mathsf{int}})$, then*

(i) $p(x_k|Y_k) = \sum_{i=0}^{1} \widehat{\nu}_k^{[i]} \mathcal{N}(\widehat{z}_k^{[i]}, \widehat{Z}_k)$, where $\widehat{\nu}_k^{[i]}$ and $\widehat{Z}_k$ are computed in Lemmas 7 and 8.

(ii) The optimal estimate $\widehat{x}_k^\varepsilon \triangleq \mathbb{E}[x_k|Y_k]$ and its EEC $P_k^\varepsilon \triangleq \mathbb{E}[(x_k - \widehat{x}_k^\varepsilon)_I^2 | Y_k]$ are the following:

$$\widehat{x}_k^\varepsilon = \sum_{i=0}^{1} \widehat{\nu}_k^{[i]} \widehat{z}_k^{[i]} \quad (21)$$

$$P_k^\varepsilon = \widehat{Z}_k + \sum_{i=0}^{1} \widehat{\nu}_k^{[i]} (\widehat{x}_k^\varepsilon - \widehat{z}_k^{[i]})_I^2. \quad (22)$$

(iii) $\mathbb{E}_{y_k}[P_k^\varepsilon] \leq \phi(\overline{P}_k^\xi, \gamma_k)$, where $\overline{P}_k^\xi = AP_{k-1}^{\mathsf{int}}A' + \overline{\nu}_k^{[0]}\overline{\nu}_k^{[1]}U_k + Q$, and $\overline{\nu}_k^{[i]}$ is computed by (19a)(19b).

4.2 Approximate optimal estimator

Lemma 9 (i) and (ii) maps $(\widehat{x}_{k-1}^{\mathsf{int}}, P_{k-1}^{\mathsf{int}}, \widehat{\nu}_{k-1}^{[i]})$ to $(\widehat{x}_k^\varepsilon, P_k^\varepsilon, \widehat{\nu}_k^{[i]})$. Denote this map by

$$(\widehat{x}_k^\varepsilon, P_k^\varepsilon, \widehat{\nu}_k^{[i]}) = \mathcal{L}(\widehat{x}_{k-1}^{\mathsf{int}}, P_{k-1}^{\mathsf{int}}, \widehat{\nu}_{k-1}^{[i]}).$$

Based on the results in Lemmas 6 and 9, we propose an algorithm (Algorithm 1) to approximately compute $\widehat{x}_k$.

Algorithm 1 Approximate state estimation algorithm for NACK systems.

Initial conditions: $\widehat{x}_0^\varepsilon = \widehat{x}_0^{\mathsf{int}} = \overline{x}_0$, $P_0^\varepsilon = P_0^{\mathsf{int}} = P_0$, $\widehat{\nu}_0^{[0]}$, and $\widehat{\nu}_0^{[1]}$.

Computation of $\widehat{x}_k^\varepsilon$ and P_k^ε, $k \in \mathbb{N}$:

Step 1 (Initial conditions): Approximate $p(x_{k-1}|Y_{k-1})$ by a Gaussian pdf $\mathcal{N}(\widehat{x}_{k-1}^{\mathsf{int}}, P_{k-1}^{\mathsf{int}})$, that is, let $p(x_{k-1}|Y_{k-1}) = \mathcal{N}(\widehat{x}_{k-1}^{\mathsf{int}}, P_{k-1}^{\mathsf{int}})$.

Step 2 (State estimate): Calculate $\widehat{x}_k^\varepsilon$ (that is, the desired approximation for $\widehat{x}_k$) by

$$(\widehat{x}_k^\varepsilon, P_k^\varepsilon, \widehat{\nu}_k^{[i]}) = \mathcal{L}(\widehat{x}_{k-1}^{\mathsf{int}}, P_{k-1}^{\mathsf{int}}, \widehat{\nu}_{k-1}^{[i]}) \quad (23)$$

and $p(x_k|Y_k)$ as in Lemma 9 (i).

Step 3: Approximate $p(x_k|Y_k)$ by a Gaussian pdf $\mathcal{N}(\widehat{x}_k^{\mathsf{int}}, P_k^{\mathsf{int}})$—which will be used as an initial condition for approximately computing $\widehat{x}_{k+1}$—via solving the following RE-based GM reduction problem: for $m \in \mathbb{R}^n, P \in \mathbb{M}_+^n$,

$$\mathscr{D}\big(p(x_k|Y_k), \mathcal{N}(\widehat{x}_k^{\mathsf{int}}, P_k^{\mathsf{int}})\big) = \min_{m,P} \mathscr{D}\big(p(x_k|Y_k), \mathcal{N}(m, P)\big).$$

Theorem 2 (Approximate optimal estimator) *Algorithm 1 can be presented in a recursive form, called the approximate optimal estimator (AOE), as follows:*

$$(\widehat{x}_k^\varepsilon, P_k^\varepsilon, \widehat{\nu}_k^{[i]}) = \mathcal{L}(\widehat{x}_{k-1}^\varepsilon, P_{k-1}^\varepsilon, \widehat{\nu}_{k-1}^{[i]}) \quad (24)$$

with $\widehat{x}_0^\varepsilon = \overline{x}_0$, $P_0^\varepsilon = P_0$, $\widehat{\nu}_0^{[0]}$, *and* $\widehat{\nu}_0^{[1]}$. *Meanwhile,*

$$\mathbb{E}_{y_k}[P_k^\varepsilon] \le \phi(\overline{P}_k^\xi, \gamma_k), \tag{25}$$

where $\overline{P}_k^\xi = AP_{k-1}^\varepsilon A' + \overline{\nu}_k^{[0]}\overline{\nu}_k^{[1]}U_k + Q$ *and* $\overline{\nu}_k^{[i]}$ *is computed by (19a) and (19b).*

Proof: By comparing with (23) and (24), it is clear that to prove (24) is to prove

$$\widehat{x}_k^{\mathsf{int}} = \widehat{x}_k^\varepsilon \text{ and } P_k^{\mathsf{int}} = P_k^\varepsilon. \tag{26}$$

(26) holds for $k=0$, due to $\widehat{x}_0^\varepsilon = \widehat{x}_0^{\mathsf{int}} = \overline{x}_0$ and $P_0^\varepsilon = P_0^{\mathsf{int}} = P_0$ in Algorithm 1. Suppose that it holds for $k = 1,\dots,n-1$. In Step 2 of Algorithm 1, $\widehat{x}_k^\varepsilon, P_k^\varepsilon, p(x_n|Y_n) = \sum_{i=0}^1 \widehat{\nu}_k^{[i]}\mathcal{N}(\widehat{z}_k^{[i]}, \widehat{Z}_k)$ can be obtained by Lemma 9 (i)(ii).

$\widehat{x}_n^{\mathsf{int}}$ and P_n^{int} in the desired Gaussian pdf $\mathcal{N}(\widehat{x}_n^{\mathsf{int}}, P_n^{\mathsf{int}})$ in Step 3 of Algorithm 1 can be calculated as follows: by letting $\{N=2,\ \alpha_i = \widehat{\nu}_n^{[i]},\ m_i = \widehat{z}_n^{[i]}$, and $P_i = \widehat{Z}_n\}$ in Lemma 6, and using (18), we have $\widehat{x}_n^{\mathsf{int}} = \sum_{i=0}^1 \widehat{\nu}_n^{[i]}\widehat{z}_n^{[i]}$ and $P_n^{\mathsf{int}} = \sum_{i=0}^1 \widehat{\nu}_n^{[i]}\big(\widehat{Z}_n + (\widehat{x}_n^\varepsilon - \widehat{z}_n^{[i]})_I^2\big)$. Comparing them with (21) and (22), we have $\widehat{x}_n^{\mathsf{int}} = \widehat{x}_n^\varepsilon$ and $P_n^{\mathsf{int}} = P_n^\varepsilon$. It follows from mathematical induction that (24) holds for $k \in \mathbb{N}$. Lemma 9 (iii) shows that $\mathbb{E}_{y_k}[P_k^\varepsilon] \le \phi(\overline{P}_k^\xi, K_k^\xi)$ holds, with $\overline{P}_k^\xi = AP_{k-1}^{\mathsf{int}}A' + \overline{\nu}_k^{[0]}\overline{\nu}_k^{[1]}U_k + Q$. (25) is proved by noting that $P_k^\varepsilon = P_k^{\mathsf{int}}$ for $k \ge 1$.

5 Performance of OE and AOE

In this section, we study the averaged estimation performance of the OE and the AOE. After providing some preliminaries, the lower and upper bounds of $\mathbb{E}_{Y_k}[P_k]$ and $\mathbb{E}_{Y_k}[P_k^\varepsilon]$ are obtained in Theorem 3. For $Y_k = \emptyset$, we define $\mathbb{E}_\emptyset[P_k] \triangleq P_k$.

Define $\overline{S}_k$ and S_k as follows:

$$\begin{aligned}\overline{S}_{k+1} &= AS_kA' + Q + \frac{1}{4}U,\\ S_k &= \phi(\overline{S}_k, \gamma_k), \text{ with } S_0 = P_0.\end{aligned} \tag{27}$$

It is easy to check that $\overline{S}_{k+1} = \overline{g}(\overline{S}_k, \gamma_k, Q + \frac{1}{4}U, R)$. Let $K_k^s = \overline{S}_kC'(C\overline{S}_kC' + R)^{-1}$. By using K_k^s, a upper bound of $\mathbb{E}_{Y_k}[P_k]$ is obtained in the following lemma.

Lemma 10 *Let* $\overline{P}_k^\ell = AP_{k-1}A' + \frac{1}{4}U_k + Q$. *Then* $\mathbb{E}_{Y_k}[P_k] \le \gamma_k h(\mathbb{E}_{Y_{k-1}}[\overline{P}_k^\ell], K_k^s) + (1-\gamma_k)\mathbb{E}_{Y_{k-1}}[\overline{P}_k^\ell]$.

Theorem 3 (Performance bounds) *The lower and upper bounds of the OE/AOE estimation performance are the following:*

$$P_k^{\mathsf{ACK}} \le \mathbb{E}_{Y_k}[P_k] \le S_k \tag{28a}$$

$$P_k^{\mathsf{ACK}} \le \mathbb{E}_{Y_k}[P_k^\varepsilon] \le S_k. \tag{28b}$$

Proof of (28a): From (17) and the result $P_k^{\mathsf{ACK}} = \widehat{M}_k$ in Theorem 1, it follows that $P_k^{\mathsf{ACK}} \le P_k$, which implies that $P_k^{\mathsf{ACK}} \le \mathbb{E}_{Y_k}[P_k]$ holds.

$\mathbb{E}_{Y_k}[P_k] \le S_k$ is proved by mathematical induction as follows. $\mathbb{E}_{y_0}[P_0] \le S_0$ holds since $P_0 = S_0$ and there is no y_0 in P_0. Suppose that $\mathbb{E}_{Y_k}[P_k] \le S_k$ holds for $k = 0,\dots,n-1$. Consider the case $k=n$ as follows. When $\gamma_n = 0$, by Lemma 10, $\mathbb{E}_{Y_n}[P_n] \le \mathbb{E}_{Y_{n-1}}[\overline{P}_n^\ell] = \mathbb{E}_{Y_{n-1}}[AP_{n-1}A' + \frac{1}{4}U_k + Q] \le AS_{n-1}A' + \frac{1}{4}U + Q = \overline{S}_n = S_n$, where the last equality is obtained by noting that $\gamma_n = 0$.
When $\gamma_n = 1$, by Lemma 10, $\mathbb{E}_{Y_n}[P_n] \le h(\mathbb{E}_{Y_{n-1}}[\overline{P}_n^\ell], K_n^s) \le h(\overline{S}_n, K_n^s) = \phi(\overline{S}_n, 1) = S_n$, where $\mathbb{E}_{Y_{n-1}}[\overline{P}_n^\ell] \le \overline{S}_n$ proved above and Lemma 2(ii)(iii) are used. The proof of (28a) is completed.

Proof of (28b): Proof of $P_k^{\mathsf{ACK}} \le \mathbb{E}_{Y_k}[P_k^\varepsilon]$: We first prove $P_k^{\mathsf{ACK}} \le P_k^\varepsilon$ by mathematical induction as follows. It holds for $k=0$ due to $P_0^{\mathsf{ACK}} = P_0^\varepsilon = P_0$. Suppose that $P_k^{\mathsf{ACK}} \le P_k^\varepsilon$ holds for $k = 0,\dots,n-1$. Then, $P_n^{\mathsf{ACK}} \overset{(a)}{=} \phi(AP_{n-1}^{\mathsf{ACK}}A' + Q, \gamma_n) \overset{(b)}{\le} \phi(AP_{n-1}^\varepsilon A' + Q, \gamma_n) \overset{(c)}{=} \widehat{Z}_n \overset{(d)}{\le} P_n^\varepsilon$, where we obtain $\overset{(a)}{=}$ by (4a) and (4b), $\overset{(b)}{\le}$ by Lemma 2 (iii), and $\overset{(c)}{=}$ by (20b), (20d), $P_{n-1}^{\mathsf{int}} = P_{n-1}^\varepsilon$ in (26), and Lemma 2 (ii); and we obtain $\overset{(d)}{\le}$ by (22). Hence, $P_n^{\mathsf{ACK}} \le P_n^\varepsilon$ holds, which implies $P_n^{\mathsf{ACK}} \le \mathbb{E}_{Y_n}[P_n^\varepsilon]$.

Proof of $\mathbb{E}_{Y_k}[P_k^\varepsilon] \le S_k$: We prove it by mathematical induction. Note that $P_0^\varepsilon = S_0 = P_0$ and that P_0^ε does not contain y_0. Thus, $\mathbb{E}_{y_0}[P_0^\varepsilon] = P_0^\varepsilon = S_0$. Suppose that it holds for $k = 0,\dots,n-1$. Consider the case $k=n$. When $\gamma_n = 0$, $\mathbb{E}_{Y_n}[P_n^\varepsilon] = \mathbb{E}_{Y_{n-1},y_n}[P_n^\varepsilon] \overset{(a)}{\le} \mathbb{E}_{Y_{n-1}}[\phi(\overline{P}_n^\xi, 0)] \overset{(b)}{=} \mathbb{E}_{Y_{n-1}}[\overline{P}_n^\xi] \overset{(c)}{\le} \overline{S}_n \overset{(d)}{=} S_n$, where $\overset{(a)}{\le}$ is obtained by (25), $\overset{(b)}{=}$ and $\overset{(d)}{=}$ are obtained by noting that $\gamma_n = 0$ in (3) and (27), and $\overset{(c)}{\le}$ is obtained as follows: $\mathbb{E}_{Y_{n-1}}[\overline{P}_n^\xi] = \mathbb{E}_{Y_{n-1}}[AP_{n-1}^\varepsilon A' + \overline{\nu}_n^{[0]}\overline{\nu}_n^{[1]}U_n + Q] \le AS_{n-1}A' + (1/4)U + Q = \overline{S}_n$. When $\gamma_n = 1$, $\mathbb{E}_{Y_n}[P_n^\varepsilon] = \mathbb{E}_{Y_{n-1},y_n}[P_n^\varepsilon] \le \mathbb{E}_{Y_{n-1}}[\phi(\overline{P}_n^\xi, 1)] = \mathbb{E}_{Y_{n-1}}[h(\overline{P}_n^\xi, K_n^\xi)] \le \mathbb{E}_{Y_{n-1}}[h(\overline{P}_n^\xi, K_n^s)] = h(\mathbb{E}_{Y_{n-1}}[\overline{P}_n^\xi], K_n^s) \le h(\overline{S}_n, K_n^s) = \phi(\overline{S}_n, 1) = S_n$. The proof of (28b) is completed.

6 Stability of OE and AOE

In this section, we establish the stability conditions for the OE and the AOE.

Definition 4 (Estimator stability) *An estimator is said to be stable in a mean sense (or stable for short), if the averaged estimation error covariance is bounded, that is,* $\sup \mathbb{E}_{\mathcal{I}_k}[P_k] < +\infty$ *for the OE, and* $\sup \mathbb{E}_{\mathcal{I}_k}[P_k^{\varepsilon}] < +\infty$ *for the AOE, where* $\mathcal{I}_k \triangleq \{Y_k, \Gamma_k\}$.

6.1 Stability relationship

Lemma 11

$$\sup \mathbb{E}_{\mathcal{I}_k}[\overline{P}_k^{\mathsf{ACK}}] < +\infty \Leftrightarrow \sup \mathbb{E}_{\mathcal{I}_k}[\overline{S}_k] < +\infty \tag{29}$$

$$\sup \mathbb{E}_{\mathcal{I}_k}[P_k^{\mathsf{ACK}}] < +\infty \Leftrightarrow \sup \mathbb{E}_{\mathcal{I}_k}[S_k] < +\infty. \tag{30}$$

Theorem 4 (Stability relationship) *For a given NACK system in (1), which satisfies Assumption 1, the following three statements are equivalent.*

(i) The OE for the NACK system is stable;
(ii) The AOE for the NACK system is stable;
(iii) The OE for the corresponding ACK system is stable.

Proof: In this proof, the subscript $\mathcal{I}_k$ of $\mathbb{E}_{\mathcal{I}_k}$ is omitted for brevity. We first prove the equivalence of (i) and (iii). By Theorem 3 and (30), we have that $\mathbb{E}[P_k^{\mathsf{ACK}}] \leq \mathbb{E}[P_k] \leq \mathbb{E}[S_k]$, and that $\mathbb{E}[S_k]$ and $\mathbb{E}[P_k^{\mathsf{ACK}}]$ have the same stability. (i) $\Rightarrow$ (iii) is proved by noting that if $\mathbb{E}[P_k]$ is stable, so is $\mathbb{E}[P_k^{\mathsf{ACK}}]$ due to $\mathbb{E}[P_k^{\mathsf{ACK}}] \leq \mathbb{E}[P_k]$; (i) $\Leftarrow$ (iii) is proved by noting that if $\mathbb{E}[P_k^{\mathsf{ACK}}]$ is stable, so is $\mathbb{E}[S_k]$ due to the same stability of $\mathbb{E}[S_k]$ and $\mathbb{E}[P_k^{\mathsf{ACK}}]$. Thus, $\mathbb{E}[P_k]$ is stable due to $\mathbb{E}[P_k] \leq \mathbb{E}[S_k]$. (i) $\Leftrightarrow$ (iii) is proved. (ii) $\Leftrightarrow$ (iii) can be proved similarly. The proof is completed.

6.2 Stability conditions for OE and AOE

To study the estimator stability for Markovian packet dropouts, a notion called "non-degeneracy" is introduced in [21, 24], which is a little bit stronger than the conventional observability and thus is called *strongly observable* in this paper. This notion is defined on diagonalizable systems, which excludes some systems with a non-trivial Jordan form, while most of real-world systems are diagonalizable and therefore the results obtained retain a great degree of generality [21]. The definition of strongly observable (that is, non-degenerate) is given as follows. In this definition, we assume that the system has already taken a diagonal form by a similarity transformation.

Definition 5 (Strongly observable)
[21, Definitions 1–5] Consider the pair (A, C) *in its diagonal standard form, that is,* $A = \mathsf{diag}(\rho_1, \ldots, \rho_n)$ *and* $C = [C_1, \ldots, C_n]$. *Let* $\rho_A \triangleq \max(|\rho_1|, \ldots, |\rho_n|)$. *A block of* (A, C) *is defined as* $A_I = \mathsf{diag}(\rho_{i_1}, \ldots, \rho_{i_j})$ *and* $C_I = [C_{i_1}, \ldots, C_{i_j}]$, *where* $1 \leq i_1 \leq i_j \leq n$ *and the index set* $I = \{i_1, \ldots, i_j\} \subseteq [1, \ldots, n]$. *A quasi-equiblock is a block in which* $|\rho_{i_1}| = \ldots = |\rho_{i_j}|$. *A pair* (A, C) *is one-step observable if* C *is full column rank. A diagonalizable system is strongly observable (also called non-degenerate) if every quasi-equiblock is one-step observable.*

Assumption 2 *A is diagonalizable and the system is strongly observable.*

Theorem 5 (Stability condition) *Consider an* $\mathcal{S}_{\mathsf{NACK}}^{\mathsf{Markov}}$ *system in (1), and suppose that Assumptions 1 and 2 hold. Then, there is a threshold value* $p_t = 1 - \rho_A^{-2}$, *where* ρ_A *is defined in Definition 5, such that*

$$\sup \mathbb{E}_{\mathcal{I}_k}[P_k] < +\infty \textit{ and } \sup \mathbb{E}_{\mathcal{I}_k}[P_k^{\varepsilon}] < +\infty, \textit{ if } p_1 > p_t;$$

for some initial P_0,

$$\sup \mathbb{E}_{\mathcal{I}_k}[P_k] = +\infty \textit{ and } \sup \mathbb{E}_{\mathcal{I}_k}[P_k^{\varepsilon}] = +\infty, \textit{ if } p_1 < p_t.$$

Proof: It is shown in [21, Theorem 8] that "under Assumptions 1 and 2, $\sup \mathbb{E}_{\mathcal{I}_k}[P_k^{\mathsf{ACK}}] < +\infty$ if $|\rho_A|^2(1 - p_1) < 1$; $\sup \mathbb{E}_{\mathcal{I}_k}[P_k^{\mathsf{ACK}}] > +\infty$, for some initial P_0, if $|\rho_A|^2(1-p_1) > 1$." Theorem 4 shows that P_k^{ACK}, P_k, and P_k^{ε} have the same stability. Therefore, we have that when $p_1 > p_t = 1-\rho_A^{-2}$ (that is, $|\rho_A|^2(1-p_1) < 1$), P_k^{ACK} is stable, so are P_k and P_k^{ε}; when $p_1 < p_t$, ($|\rho_A|^2(1-p_1) > 1$), P_k^{ACK} is unstable, so are P_k and P_k^{ε}. The proof is completed.

7 Simulation Examples

In this section, some numerical examples are presented to illustrate the main results of this paper. Consider the system in (1) with the following parameters:

$$A = \begin{bmatrix} \rho_A & 0 \\ 0 & 1.005 \end{bmatrix}, B = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, C = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}, Q = R = \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix},$$

where A has been presented in a diagonal form and $\rho_A > 1.005$. The following simulations are performed under bounded control inputs $u_k = 1 + \exp(-k/10)$.

Stability: Figures 2 and 3 show the relationship between the expected EEC of the OE/AOE and the recovery/failure rate. Let ρ_A take different values $\{1.2910, 1.8257\}$, and the corresponding threshold value $q_t = 1 - \rho_A^{-2}$ are $\{0.4, 0.7\}$. Fig. 2 shows that $\mathbb{E}_{\mathcal{I}_{30}}[P_{30}]$ and $\mathbb{E}_{\mathcal{I}_{30}}[P_{30}^{\varepsilon}]$ enlarge and tend to unstable as p_1 deceases and approaches the threshold value q_t. Fig. 3

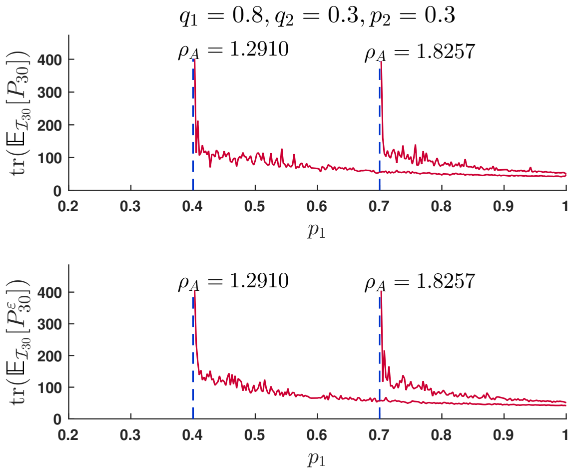

Fig. 2. The relationship between the observation recovery rate p_1 and $\mathrm{tr}(\mathbb{E}_{\mathcal{I}_{30}}[P_{30}])$, $\mathrm{tr}(\mathbb{E}_{\mathcal{I}_{30}}[P^{\varepsilon}_{30}])$.

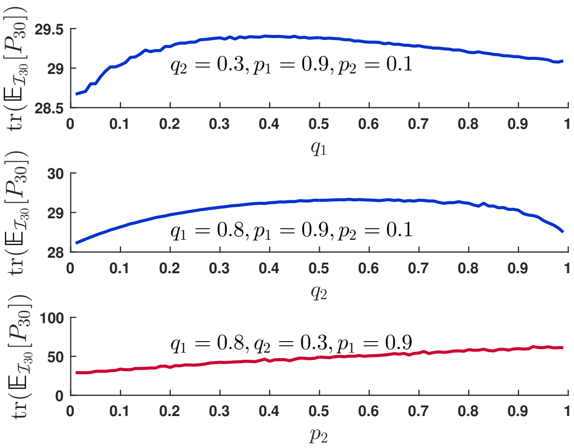

Fig. 3. The relationship between $\mathrm{tr}(\mathbb{E}_{\mathcal{I}_{30}}[P_{30}])$ and the control packet recovery rate q_1/the control packet failure rate q_2/the observation failure rate p_2.

shows that other parameters q_1, q_2, p_2 do not affect the stability of the OE, and the similar graphs for the AOE are not presented for saving space. These phenomena on the estimator stability agree with the statements of Theorems 4 and 5.

Performance: Figure 4 shows that the expected estimation performance of the OE $\mathbb{E}_{Y_k}[P_k]$ and the AOE $\mathbb{E}_{Y_k}[P^{\varepsilon}_k]$ lie between the upper bound $\mathbb{E}_{Y_k}[S_k]$ and the lower one $\mathbb{E}_{Y_k}[P^{\mathsf{ACK}}_k]$, as stated in Theorems 3. In Fig. 5, $\mathbb{E}_{\mathcal{I}_k}[P^{\varepsilon}_k]$ lies between $\sup \mathbb{E}_{\mathcal{I}_k}[S_k]$ and $\inf \mathbb{E}_{\mathcal{I}_k}[P^{\mathsf{ACK}}_k]$. $\mathbb{E}_{\mathcal{I}_k}[P_k]$ is not presented due to time-consuming computation, but according to Theorem 3, it must also lie between $\sup \mathbb{E}_{\mathcal{I}_k}[S_k]$ and $\inf \mathbb{E}_{\mathcal{I}_k}[P^{\mathsf{ACK}}_k]$.

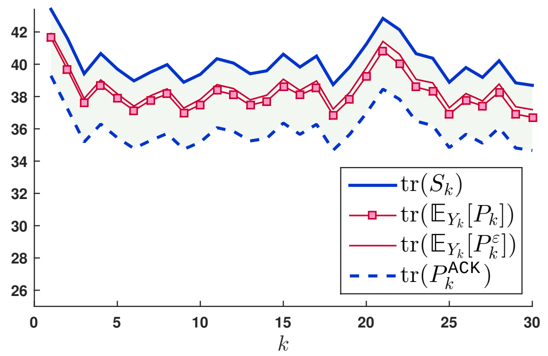

Fig. 4. Lower and upper bounds of $\mathrm{tr}(\mathbb{E}_{Y_k}[P_k])$ and $\mathrm{tr}(\mathbb{E}_{Y_k}[P^{\varepsilon}_k])$

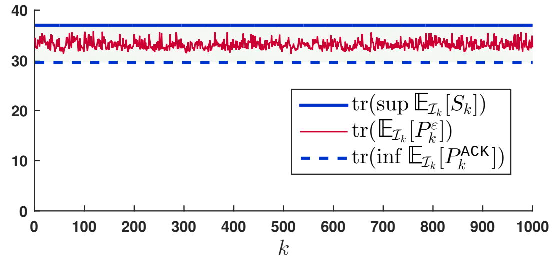

Fig. 5. Lower and upper bounds of $\mathrm{tr}(\mathbb{E}_{\mathcal{I}_k}[P^{\varepsilon}_k])$

8 Conclusion

In this paper, we have investigated state estimation for systems with both control and observation packet dropouts. We have obtained the OE for NACK systems with Markovian packet dropouts, and developed an computationally efficient AOE, with the estimation performance close to the OE. We have shown that the stability of the OE and the AOE for strongly observable NACK systems only depends on the observation packet recovery rate p_1. There are two future research topics: (i) To improve the estimation performance for NACK systems, we may consider using smart devices such as smart sensors [16] and network relays [11], or employing advanced estimation techniques such as event-based estimation methods [17] and redundant communication channels design schemes [27, 38]. (ii) Besides NACK and ACK systems, it is also meaningful to investigate a system with partial acknowledgments [14]. An interesting topic is how to recursively encode all (or most of) the historical acknowledgments into a packet at the actuator side. If it can be achieved, then the estimator will acquire more information on acknowledgments by decoding the packet after receiving it, and the estimation performance may be improved.

Appendix

Proof of Lemma 2: Part (i) can be readily proved either by following the same derivation process as Lemma 1 of [26], or by letting $A = I$ and $\lambda = 1$ in [26, Eq. (28)] and using the result [26, Lemma 1 (a)].

Part (ii) is a commonly used equality in Kalman filer, which can be easily proved by some algebraic manipulations, or by the results on the P_k in [28, p. 144].

Proof of (iii): It is easy to check that $h(X,K) \le h(Y,K)$ and $\phi(X,0) \le \phi(Y,0)$ hold for $X \le Y$. When $\gamma = 1$, $\phi(X,1) = h(X,K_X) \le h(X,K_Y) \le h(Y,K_Y) = \phi(Y,1)$. $g(X,\gamma) \le g(Y,\gamma)$ holds by noting that $g(Y,\gamma) - g(X,\gamma) = A(\phi(Y,\gamma) - \phi(X,\gamma))A' \ge 0$.

The result of part (iv) is trivial.

Proof of Lemma 3: Proof of (i): (6) is obtained by the property of Markov processes that whenever $\nu_n = j_1$ is known, $p(\nu_{n+1} = j_2|\nu_n = j_1, \Upsilon_{n-1}^{[i]})$ is determined by $p(\nu_{n+1} = j_2|\nu_n = j_1)$ and is independent of $\{\nu_{n-1}, \dots, \nu_1\}$. Similarly, by the property of Markov processes, whenever ν_n in $\Upsilon_n^{[i]}$ is known, $p(\nu_{n+1} = j_2|\Upsilon_n^{[i]})$ is independent of Y_n, which proves (7).

Proof of (ii): For $i \in \mathbb{I}_1^{2^{n-1}}$, by using (5) and Lemma 3 (i), $p(\nu_{n+1} = 0|\Upsilon_n^{[i]}) = p(\nu_{n+1} = 0|\nu_n = 0, \Upsilon_{n-1}^{[i]}) = p(\nu_{n+1} = 0|\nu_n = 0) = 1 - q_1$. Other cases can be proved similarly.

Proof of Lemma 4: We prove this lemma by mathematical induction. Consider the case $k = 0$. Note that $Y_0 = \emptyset$ and $\Upsilon_0^{[i]} = \emptyset$. $p(x_0) = \mathcal{N}(\overline{x}_0, P_0) = \mathcal{N}(\widehat{m}_0^{[1]}, \widehat{M}_0)$. (10) holds for $k = 0$.

By the definition of $\Upsilon_k^{[i]}$, $\Upsilon_1^{[1]} = (\nu_1 = 0)$ and $\Upsilon_1^{[2]} = (\nu_1 = 1)$. Consider the case $\nu_1 = 0$. Using Lemma 1 and letting $\{X = x_1, Z = x_0, u = \nu_1 u_1 = 0, W = \omega_1\}$, and using (2a), we can obtain $p(x_1|\Upsilon_1^{[1]}) = \mathcal{N}(A\overline{x}_0, AP_0A' + Q) = \mathcal{N}(\overline{m}_1^{[1]}, \overline{M}_1)$. Similarly, we have $p(x_1|\Upsilon_1^{[2]}) = \mathcal{N}(A\overline{x}_0 + Bu_1, AP_0A' + Q) = \mathcal{N}(\overline{m}_1^{[2]}, \overline{M}_1)$. Thus, (11) holds for $k = 0$.

Suppose that Lemma 4 holds for $k = 0, \dots, n-1$. Consider the case $k = n$ as follows. When $\gamma_n = 0$, no y_n is available, and thus $Y_n = Y_{n-1}$. $p(x_n|Y_n, \Upsilon_n^{[i]}) = p(x_n|Y_{n-1}, \Upsilon_n^{[i]}) = \mathcal{N}(\overline{m}_n^{[i]}, \overline{M}_n) = \mathcal{N}(\widehat{m}_n^{[i]}, \widehat{M}_n)$ with $\gamma_n = 0$, which proves (10) with $\gamma_n = 0$.

When $\gamma_n = 1$, $p(x_n|Y_n, \Upsilon_n^{[i]}) = p(x_n|y_n, Y_{n-1}, \Upsilon_n^{[i]})$. By viewing $\{Y, X, V\}$ in Lemma 1 as $\{y_n, x_n, \upsilon_n\}$ respectively, letting $p(X)$ in (2a) be $p(X) = p(x_n|Y_{n-1}, \Upsilon_n^{[i]}) = \mathcal{N}(\overline{m}_n^{[i]}, \overline{M}_n)$, and then using (2b), $p(x_n|y_n, Y_{n-1}, \Upsilon_n^{[i]}) = \mathcal{N}(\overline{m}_n^{[i]} + K_n(y_n - C\overline{m}_n^{[i]}), (I - K_nC)\overline{M}_n) = \mathcal{N}(\widehat{m}_n^{[i]}, \widehat{M}_n)$, which proves (10) with $\gamma_n = 1$.

By (5), $\Upsilon_{n+1}^{[i]} = \{\nu_{n+1} = 0, \Upsilon_n^{[i]}\}$ with $1 \le i \le 2^n$. By viewing $x_{n+1} = Ax_n + \omega_{n+1}$ as $X = AZ + Bu + W$ in Lemma 1 with $u = 0$, and letting $p(Z) = p(x_n|\Upsilon_n^{[i]}, Y_n) = \mathcal{N}(\widehat{m}_n^{[i]}, \widehat{M}_n)$, and then using (2a), we have $p(x_{n+1}|\nu_{n+1} = 0, \Upsilon_n^{[i]}, Y_n) = \mathcal{N}(A\widehat{m}_n^{[i]}, A\widehat{M}_nA' + Q) = \mathcal{N}(\overline{m}_{n+1}^{[i]}, \overline{M}_{n+1}) = p(x_{n+1}|\Upsilon_{n+1}^{[i]}, Y_n)$ for $1 \le i \le 2^n$. Similarly, for $2^n + 1 \le i \le 2^{n+1}$, we have $p(x_{n+1}|\Upsilon_{n+1}^{[i]}, Y_n) = \mathcal{N}(\overline{m}_{n+1}^{[i]}, \overline{M}_{n+1})$. Therefore, (11) holds for $k = n$. The proof is completed.

Proof of Lemma 5: When $\gamma_n = 0$, $Y_n = Y_{n-1}$. $\widehat{\alpha}_n^{[i]} = p(\Upsilon_n^{[i]}|Y_n) = p(\Upsilon_n^{[i]}|Y_{n-1}) = \overline{\alpha}_n^{[i]}$. When $\gamma_n = 1$, by using Bayesian formula, it is easy to obtain that $p(\Upsilon_n^{[i]}|Y_n) = p(y_n|\Upsilon_n^{[i]}, Y_{n-1})\frac{p(\Upsilon_n^{[i]}|Y_{n-1})}{\sum_{i=1}^{2^n} p(y_n|\Upsilon_n^{[i]}, Y_{n-1})p(\Upsilon_n^{[i]}|Y_{n-1})} = \widehat{\alpha}_n^{[i]}$, which proves (14). For $1 \le i \le 2^n$, by (5), $\Upsilon_{n+1}^{[i]} = \{\nu_{n+1} = 0, \Upsilon_n^{[i]}\}$.

$$
\begin{aligned}
p(\Upsilon_{n+1}^{[i]}|Y_n) &= p(\nu_{n+1} = 0, \Upsilon_n^{[i]}|Y_n) \\
&= p(\nu_{n+1} = 0|\Upsilon_n^{[i]}, Y_n)p(\Upsilon_n^{[i]}|Y_n) \\
&= \begin{cases} \widehat{\alpha}_n^{[i]}(1 - q_1), & i \in \mathbb{I}_1^{2^{n-1}} \\ \widehat{\alpha}_n^{[i]} q_2, & i \in \mathbb{I}_{2^{n-1}+1}^{2 \cdot 2^{n-1}} \end{cases}
\end{aligned}
$$

where the last equality are obtained by Lemma 3 (ii). The remaining two cases in (15) can be proved similarly by (5) and Lemma 3. The proof is completed.

Proof of Lemma 6: According to the definition of relative entropy, $\mathscr{D}\big(p, \mathcal{N}(m,P)\big) = \int p \log(p/\mathcal{N}(m,P))\mathrm{d}x = \int p(x)\log p(x)\mathrm{d}x + n\log 2\pi + \log\det(P) + \Psi$, where it is easy to obtain that $\Psi = \sum_{i=1}^N \alpha_i \mathrm{tr}(P^{-1}(P_i + (m_i - m)_I^2))$. By solving $\frac{\partial \Psi}{\partial m} = 0$, we obtain $m^* = \sum_{i=1}^N \alpha_i m_i$, which is a minimum point, as $\frac{\partial^2 \Psi}{\partial m^2} = 2\mathrm{tr}P^{-1} > 0$.

By letting $m = m^*$, $\Psi = \sum_{i=1}^N \alpha_i \mathrm{tr}(P^{-1}(P_i + (m_i - m^*)_I^2)) = \mathrm{tr}(P^{-1}P^*)$, where P^* is given in (18). Let $f(P) = \log\det(P) + \Psi = \log\det(P) + \mathrm{tr}(P^{-1}P^*)$. Then, $\frac{\partial \mathscr{D}\left(p, \mathcal{N}(m,P)\right)}{\partial P} = \frac{\partial f(P)}{\partial P} = P^{-1} - P^{-2}P^*$. Solving $\frac{\partial \mathscr{D}}{\partial P} = 0$ with $P > 0$ yields $P = P^*$, which is a minimum point, since $\frac{\partial^2 \mathscr{D}}{\partial P^2}\big|_{P=P^*} = \frac{\partial^2 f(P)}{\partial P^2}\big|_{P=P^*} = (P^*)^{-2} > 0$. The proof is completed.

Proof of Lemma 7: By definition, $\overline{\nu}_k^{[0]} = p(\nu_k = 0|Y_{k-1}) = \sum_{i=0}^{1} p(\nu_k = 0|\nu_{k-1} = i, Y_{k-1})p(\nu_{k-1} = i|Y_{k-1}) = (1-q_1)\widehat{\nu}_{k-1}^{[0]} + q_2\widehat{\nu}_{k-1}^{[1]}$, which proves (19a). $\overline{\nu}_k^{[1]} = p(\nu_k = 1|Y_{k-1}) = 1 - p(\nu_k = 0|Y_{k-1}) = q_1\widehat{\nu}_{k-1}^{[0]} + (1-q_2)\widehat{\nu}_{k-1}^{[1]}$, which proves (19b).

When $\gamma_k = 0$, no observation y_k is available and thus $Y_k = Y_{k-1}$. $\widehat{\nu}_k^{[i]} = p(\nu_k = i|Y_k) = p(\nu_k = i|Y_{k-1}) = \overline{\nu}_k^{[i]}$. When $\gamma_k = 1$, by Bayesian formula, $\widehat{\nu}_k^{[i]} = p(\nu_k = i|Y_k) = p(\nu_k = i|y_k, Y_{k-1}) = p(y_k|\nu_k = i, Y_{k-1})p(\nu_k = i|Y_{k-1})/p(y_k|Y_{k-1}) = \varpi_k^{[i]}\overline{\nu}_k^{[i]}/\sum_{i=0}^{1}\varpi_k^{[i]}\overline{\nu}_k^{[i]}$, which proves (19c).

Proof of Lemma 8: From $p(x_{k-1}|Y_{k-1})$ and by using (2a), we have $p(x_k|\nu_k = i, Y_{k-1}) = \mathcal{N}(\overline{z}_k^{[i]}, \overline{Z}_k)$.

When $\gamma_k = 0$, $Y_k = Y_{k-1}$. Then, we have $\widehat{z}_k^{[i]} = \overline{z}_k^{[i]}$ and $\widehat{Z}_k = \overline{Z}_k$, which is conformed with (20c) and (20d), where $K_k^{\varepsilon} = 0$ when $\gamma_k = 0$.

When $\gamma_k = 1$, from the pdf $p(x_k|\nu_k = i, Y_{k-1})$ obtained above, by using (2b), we obtain the measurement-update pdf $p(x_k|\nu_k = i, y_k, Y_{k-1}) = \mathcal{N}(\widehat{z}_k^{[i]}, \widehat{Z}_k)$, where $\widehat{z}_k^{[i]}$ and $\widehat{Z}_k$ are computed in the same way as (20c) and (20d).

Proof of Lemma 9: Proof of (i): From $\widehat{\nu}_k^{[i]} \triangleq p(\nu_k = i|Y_k)$ and $p(x_k|\nu_k = i, Y_k)$ computed above, $p(x_k|Y_k) = \sum_{i=0}^{1} p(\nu_k = i|Y_k)p(x_k|\nu_k = i, Y_k) = \sum_{i=0}^{1}\widehat{\nu}_k^{[i]}\mathcal{N}(\widehat{z}_k^{[i]}, \widehat{Z}_k)$, which proves (i).

Proof of (ii): It is a well-known result in [1, p. 213], that for a Gaussian mixture $p(x_k|Y_k) = \sum_{i=0}^{1}\widehat{\nu}_k^{[i]}\mathcal{N}(\widehat{z}_k^{[i]}, \widehat{Z}_k)$, the mean $\widehat{x}_k^{\varepsilon}$ and the covariance P_k^{ε} can be computed as (21) and (22).

Proof of (iii): Construct an estimate $\widehat{x}_k^{\xi}$ as follows:

$$\begin{aligned}\overline{x}_k^{\xi} &= A\widehat{x}_{k-1}^{\mathsf{int}} + \overline{\nu}_k^{[1]}Bu_k\\ \widehat{x}_k^{\xi} &= \overline{x}_k^{\xi} + \gamma_k K_k^{\xi}(y_k - C\overline{x}_k^{\xi}),\end{aligned}$$

where $K_k^{\xi} = \overline{P}_k^{\xi}C'(C\overline{P}_k^{\xi}C' + R)^{-1}$. The estimation error $x_k - \overline{x}_k^{\xi} = A(x_{k-1} - \widehat{x}_{k-1}^{\mathsf{int}}) + (\nu_k - \overline{\nu}_k^{[1]})Bu_k + \omega_k$. We first compute $\Delta_i \triangleq \mathbb{E}[(x_k - \overline{x}_k^{\xi})_I^2|\nu_k = i, Y_{k-1}]$ as follows: For $i = 0$, $\Delta_0 = \Delta_A + \Delta_B + \Delta_B' + (\overline{\nu}_k^{[1]})^2Bu_ku_k'B' + Q$, where $\Delta_A \triangleq A\mathbb{E}[(x_{k-1} - \widehat{x}_{k-1}^{\mathsf{int}})_I^2|Y_{k-1}]A'$ and $\Delta_B \triangleq \mathbb{E}[(x_{k-1} - \widehat{x}_{k-1}^{\mathsf{int}})(\overline{\nu}_k^{[1]}Bu_k)'|Y_{k-1}]$. Since $\overline{\nu}_k^{[1]}$ is a function of Y_{k-1} and does not contain x_{k-1}, we have $\Delta_B = \mathbb{E}[(x_{k-1} - \widehat{x}_{k-1}^{\mathsf{int}})|Y_{k-1}](\overline{\nu}_k^{[1]}Bu_k)' = 0$. Consequently, $\Delta_0 = \Delta_A + (\overline{\nu}_k^{[1]})^2U_k + Q$. Similarly, $\Delta_1 = \Delta_A + (\overline{\nu}_k^{[0]})^2U_k + Q$. $\overline{P}_k^{\xi} \triangleq \mathbb{E}_{\nu_k}[\mathbb{E}[(x_k - \overline{x}_k^{\xi})_I^2|\nu_k, Y_{k-1}]] = \Delta_A + \overline{\nu}_k^{[0]}\overline{\nu}_k^{[1]}U_k + Q$.

When $\gamma_k = 0$, $\widehat{z}_k^{[i]} = \overline{z}_k^{[i]}$, $\widehat{\nu}_k^{[i]} = \overline{\nu}_k^{[i]}$, $\widehat{Z}_k = \overline{Z}_k$, and $K_k^{\varepsilon} = 0$. By (20b), (21), and (22), we have $P_k^{\varepsilon} = AP_{k-1}^{\mathsf{int}}A' + \overline{\nu}_k^{[0]}\overline{\nu}_k^{[1]}U_k + Q = \overline{P}_k^{\xi} = \phi(\overline{P}_k^{\xi}, 0)$, which implies $\mathbb{E}_{y_k}[P_k^{\varepsilon}] \leq \mathbb{E}_{y_k}[\phi(\overline{P}_k^{\xi}, 0)] = \phi(\overline{P}_k^{\xi}, 0)$, where the last equality is obtained by noting that there is no y_k in $\phi(\overline{P}_k^{\xi}, 0)$.

When $\gamma_k = 1$, $x_k - \widehat{x}_k^{\xi} = (I - K_k^{\xi}C)(x_k - \overline{x}_k^{\xi}) + K_k^{\xi}\upsilon_k$. $P_k^{\xi} \triangleq \mathbb{E}[(x_k - \widehat{x}_k^{\xi})_I^2|Y_{k-1}] = (I - K_k^{\xi}C)\overline{P}_k^{\xi}(I - K_k^{\xi}C)' + K_k^{\xi}RK_k^{\xi\prime} = h(\overline{P}_k^{\xi}, K_k^{\xi}) = \phi(\overline{P}_k^{\xi}, 1)$. For $\gamma_k = 1$, $\mathbb{E}_{y_k}[P_k^{\varepsilon}] \overset{(a)}{=} \mathbb{E}_{y_k}[\mathbb{E}[(x_k - \widehat{x}_k^{\varepsilon})_I^2|Y_k]] \overset{(b)}{\leq} \mathbb{E}_{y_k}[\mathbb{E}[(x_k - \widehat{x}_k^{\xi})_I^2|Y_k]] \overset{(c)}{=} \mathbb{E}[(x_k - \widehat{x}_k^{\xi})_I^2|Y_{k-1}] = P_k^{\xi} = \phi(\overline{P}_k^{\xi}, 1)$, where we obtain $\overset{(a)}{=}$ by the definition of P_k^{ε} in Lemma 9 (ii), $\overset{(b)}{\leq}$ by noting that $\mathbb{E}[(x_k - \widehat{x}_k^{\varepsilon})_I^2|Y_k] \leq \mathbb{E}[(x_k - \widehat{x}_k^{\xi})_I^2|Y_k]$, $\overset{(c)}{=}$ by using the property that $\mathbb{E}_{y_k}[\mathbb{E}[\cdot|Y_{k-1}, y_k]] = \mathbb{E}[\cdot|Y_{k-1}]$.

It can be concluded from the above that $\mathbb{E}_{y_k}[P_k^{\varepsilon}] \leq \phi(\overline{P}_k^{\xi}, \gamma_k)$ for $\gamma_k = 0$ and 1, which proves (iii).

Proof of Lemma 10: Construct an estimate $\widehat{x}_k^{\ell}$:

$$\begin{aligned}\overline{x}_k^{\ell} &= A\widehat{x}_{k-1} + (1/2)Bu_k\\ \widehat{x}_k^{\ell} &= \overline{x}_k^{\ell} + \gamma_k K_k^{s}(y_k - C\overline{x}_k^{\ell}).\end{aligned}$$

The estimation error $x_k - \overline{x}_k^{\ell} = A(x_{k-1} - \widehat{x}_{k-1}) + (\nu_k - \frac{1}{2})Bu_k + \omega_k$. We first compute $\Delta_i^{\ell} \triangleq \mathbb{E}[(x_k - \overline{x}_k^{\ell})_I^2|\nu_k = i, Y_{k-1}]$ as follows: For $i = 0$, $\Delta_0^{\ell} = \Delta_A^{\ell} + \Delta_B^{\ell} + (\Delta_B^{\ell})' + (\frac{1}{2})^2U_k + Q$, where $\Delta_A^{\ell} \triangleq A\mathbb{E}[(x_{k-1} - \widehat{x}_{k-1})_I^2|Y_{k-1}]A'$ and $\Delta_B^{\ell} \triangleq \mathbb{E}[(x_{k-1} - \widehat{x}_{k-1})(1/2Bu_k)'|Y_{k-1}] = 0$. Similar to the proof of Lemma 9 (iii), we have $\Delta_1^{\ell} = \Delta_A^{\ell} + (\frac{1}{2})^2U_k + Q = \Delta_0^{\ell}$ and then obtain $\overline{P}_k^{\ell} \triangleq \mathbb{E}_{\nu_k}[\mathbb{E}[(x_k - \overline{x}_k^{\ell})_I^2|\nu_k, Y_{k-1}]] = \Delta_A^{\ell} + (1/2)^2U_k + Q$.

When $\gamma_k = 0$, $\widehat{x}_k^{\ell} = \overline{x}_k^{\ell}$. Thus, $P_k^{\ell} \triangleq \mathbb{E}[(x_k - \widehat{x}_k^{\ell})_I^2|Y_{k-1}] = \overline{P}_k^{\ell}$. Then, we have $\mathbb{E}_{Y_{k-1}}[P_k^{\ell}] = \mathbb{E}_{Y_{k-1}}[\overline{P}_k^{\ell}]$.

When $\gamma_k = 1$, $x_k - \widehat{x}_k^{\ell} = (I - K_k^{s}C)(x_k - \overline{x}_k^{\ell}) + K_k^{s}\upsilon_k$. $P_k^{\ell} = (I - K_k^{s}C)\mathbb{E}[(x_k - \overline{x}_k^{\ell})_I^2|Y_{k-1}](I - K_k^{s}C)' + \mathbb{E}[(K_k^{s}\upsilon_k)_I^2|Y_{k-1}] = (I - K_k^{s}C)\overline{P}_k^{\ell}(I - K_k^{s}C)' + K_k^{s}RK_k^{s\prime}$. Then, we have $\mathbb{E}_{Y_{k-1}}[P_k^{\ell}] = h(\mathbb{E}_{Y_{k-1}}[\overline{P}_k^{\ell}], K_k^{s})$.

$\mathbb{E}_{Y_k}[P_k] = \mathbb{E}_{Y_k}[\mathbb{E}[(x_k - \widehat{x}_k)_I^2|Y_k]] \leq \mathbb{E}_{Y_{k-1}, y_k}[\mathbb{E}[(x_k - \widehat{x}_k^{\ell})_I^2|Y_{k-1}, y_k]] = \mathbb{E}_{Y_{k-1}}[\mathbb{E}[(x_k - \widehat{x}_k^{\ell})_I^2|Y_{k-1}]] = \mathbb{E}_{Y_{k-1}}[P_k^{\ell}] = \gamma_k h(\mathbb{E}_{Y_{k-1}}[\overline{P}_k^{\ell}], K_k^{s}) + (1 - \gamma_k)\mathbb{E}_{Y_{k-1}}[\overline{P}_k^{\ell}]$. The proof is completed.

Proof of Lemma 11: Proof of (29): For $\{Q, R, Q + \frac{1}{4}U, \overline{S}_1, \overline{P}_1^{\mathsf{ACK}}\}$, there always exist positive real numbers $\underline{\epsilon}$ and $\overline{\epsilon}$, such that 1) $\underline{\epsilon} I_n$ and $\overline{\epsilon} I_n$ are lower and upper bounds of matrices $\{Q, R, Q + \frac{1}{4}U, \overline{S}_1, \overline{P}_1^{\mathsf{ACK}}\}$, respectively; and 2) $\underline{\epsilon} I_m < R < \overline{\epsilon} I_m$.

For notional brevity, the subscript $\mathcal{I}_k$ of $\mathbb{E}_{\mathcal{I}_k}$ and the subscript of I_n and I_m are omitted, and assume the identity matrix I has an appropriate dimension for matrix manipulations. Define three sequences as follows:

$$\begin{aligned} \underline{X}_{k+1} &= \overline{g}(\underline{X}_k, \gamma_k, \underline{\epsilon} I, \underline{\epsilon} I) \text{ with } \underline{X}_1 = \underline{\epsilon} I \\ \overline{X}_{k+1} &= \overline{g}(\overline{X}_k, \gamma_k, \overline{\epsilon} I, \overline{\epsilon} I) \text{ with } \overline{X}_1 = \overline{\epsilon} I \\ T_{k+1} &= \overline{g}(T_k, \gamma_k, I, I) \text{ with } T_1 = I. \end{aligned}$$

We will prove that the following facts hold.

(i) $\underline{X}_k \le \overline{P}_k^{\mathsf{ACK}} \le \overline{X}_k$, and $\underline{X}_k \le \overline{S}_k \le \overline{X}_k$;
(ii) $\underline{X}_k = \underline{\epsilon} T_k$ and $\overline{X}_k = \overline{\epsilon} T_k$.

Proof of (i): Clearly, $\underline{X}_1 \le \overline{P}_1^{\mathsf{ACK}} \le \overline{X}_1$ holds. If $\underline{X}_k \le \overline{P}_k^{\mathsf{ACK}} \le \overline{X}_k$ holds, then, by Lemma 2 (iv), $\overline{X}_{k+1} = \overline{g}(\overline{X}_k, \gamma_k, \overline{\epsilon} I, \overline{\epsilon} I) \ge \overline{P}_{k+1}^{\mathsf{ACK}} = \overline{g}(\overline{P}_k^{\mathsf{ACK}}, \gamma_k, Q, R) \ge \overline{g}(\underline{X}_k, \gamma_k, \underline{\epsilon} I, \underline{\epsilon} I) = \underline{X}_{k+1}$. Similarly, it is easy to prove that $\underline{X}_k \le \overline{S}_k \le \overline{X}_k$.

Proof of (ii): Note that $\underline{X}_1 = \underline{\epsilon} I = \underline{\epsilon} T_1$. Suppose that $\underline{X}_k = \underline{\epsilon} T_k$ holds. Then, $\underline{X}_{k+1} = \overline{g}(\underline{X}_k, \gamma_k, \underline{\epsilon} I, \underline{\epsilon} I) = \overline{g}(\underline{\epsilon} T_k, \gamma_k, \underline{\epsilon} I, \underline{\epsilon} I) = \underline{\epsilon}\overline{g}(T_k, \gamma_k, I, I) = \underline{\epsilon} T_{k+1}$. Similarly, we can show that $\overline{X}_k = \overline{\epsilon} T_k$ holds.

From the facts (i)(ii), we have $\underline{\epsilon} T_k = \underline{X}_k \le \overline{P}_k^{\mathsf{ACK}} \le \overline{X}_k = \overline{\epsilon} T_k$, and $\underline{\epsilon} T_k \le \overline{S}_k \le \overline{\epsilon} T_k$. Thus, it is clear that $\sup \mathbb{E}[\overline{P}_k^{\mathsf{ACK}}] < +\infty$ is equivalent to $\sup \mathbb{E}[\overline{S}_k] < +\infty$. (29) is proved.

Proof of (30): From (4b), it is clear that $P_k^{\mathsf{ACK}} = \phi(\overline{P}_k^{\mathsf{ACK}}, \gamma_k) \le \overline{P}_k^{\mathsf{ACK}}$, which means that if $\sup \mathbb{E}[\overline{P}_k^{\mathsf{ACK}}]$ is bounded, so is $\sup \mathbb{E}[P_k^{\mathsf{ACK}}]$. From (4a), we have if $\sup \mathbb{E}[P_k^{\mathsf{ACK}}]$ is bounded, so is $\sup \mathbb{E}[P_{k+1}^{\mathsf{ACK}}]$. Therefore, $\sup \mathbb{E}[P_k^{\mathsf{ACK}}] < +\infty \Leftrightarrow \sup \mathbb{E}[\overline{P}_k^{\mathsf{ACK}}] < +\infty$ holds. The equivalence $\sup \mathbb{E}[S_k] < +\infty \Leftrightarrow \sup \mathbb{E}[\overline{S}_k] < +\infty$ can be proved similarly. By (29), we have $\sup \mathbb{E}[P_k^{\mathsf{ACK}}] < +\infty \Leftrightarrow \sup \mathbb{E}[S_k] < +\infty$. The proof of (30) is completed.

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Englewood Cliffs: Prentice-Hall, 1979.

[2] G. Bernieri, E. E. Miciolino, F. Pascucci, and R. Setola. Monitoring system reaction in cyber-physical testbed under cyber-attacks. *Computers & Electrical Engineering*, 59:86–98, 2017.

[3] R. Blind and F. Allgöwer. Estimating the fates of the control packets for networked control systems with loss of control and measurement packets. In *Proc. the 48th IEEE CDC/CCC*, pages 2687–2692, 2009.

[4] O. L. V. Costa. Linear minimum mean square error estimation for discrete-time Markovian jump linear systems. *IEEE Trans. Automat. Control*, 39(8):1685–1689, 1994.

[5] H. Dong, Z. Wang, J. Lam, and H. Gao. Distributed filtering in sensor networks with randomly occurring saturations and successive packet dropouts. *Int. J. Robust Nonlin. Control*, 24(12):1743–1759, 2014.

[6] B. A. Forouzan and S. C. Fegan. *TCP/IP Protocol Suite*. McGraw-Hill Higher Education, 2002.

[7] M. Hadidi and S. Schwartz. Linear recursive state estimators under uncertain observations. *IEEE Trans. Autom. Control*, 24(6):944–948, 1979.

[8] J. R. Hershey and P. A. Olsen. Approximating the Kullback Leibler divergence between Gaussian mixture models. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE Int. Conf.*, volume 4, pages IV–317. IEEE, 2007.

[9] M. Huang and S. Dey. Stability of Kalman filtering with Markovian packet losses. *Automatica*, 43(4):598–607, 2007.

[10] S. Kullback. *Information Theory and Statistics*. Courier Corporation, 1997.

[11] A. S. Leong and D. E. Quevedo. Kalman filtering with relays over wireless fading channels. *IEEE Trans. Autom. Control*, 61(6):1643–1648, 2016.

[12] H. Lin, J. Lam, M. Z. Q. Chen, Z. Shu, and Z.-G. Wu. Interacting multiple model estimator for networked control systems: Stability, convergence, and performance. *IEEE Trans. Autom. Control*, 64(3):928–943, 2019.

[13] H. Lin, H. Su, M. Z. Q. Chen, Z. Shu, R. Lu, and Z.-G. Wu. On stability and convergence of optimal estimation for networked control systems with dual packet losses without acknowledgment. *Automatica*, 90:81–90, 2018.

[14] H. Lin, H. Su, P. Shi, R. Lu, and Z.-G. Wu. Estimation and LQG control over unreliable network with acknowledgment randomly lost. *IEEE Trans. Cybern.*, 47(12):4074–4085, 2017.

[15] H. Lin, H. Su, P. Shi, Z. Shu, R. Lu, and Z.-G. Wu. Optimal estimation and control for lossy network: stability, convergence, and performance. *IEEE Trans. Autom. Control*, 62(9):4564–4579, 2017.

[16] H. Lin, H. Su, Z. Shu, Z.-G. Wu, and Y. Xu. Optimal estimation in UDP-like networked control systems with intermittent inputs: Stability analysis and suboptimal filter design. *IEEE Trans. Autom. Control*, 61(7):1794–1809, 2016.

[17] Q. Liu, Z. Wang, X. He, and D.-H. Zhou. Event-based recursive distributed filtering over wireless sensor networks. *IEEE Trans. Autom. Control*, 60(9):2470–2475, 2015.

[18] W. Liu, K. Yu, X. Tan, and X. Bu. State estimation for networked systems with Markovian communication constraints and multiple packet dropouts. *J. Franklin Inst.*, 355(2):878–901, 2018.

[19] X. F. Liu, M. R. Shahriar, S. N. Al Sunny, M. C. Leu, and L. Hu. Cyber-physical manufacturing cloud: Architecture, virtualization, communication, and testbed. *J. Manufacturing Syst.*, 43:352–364, 2017.

[20] P. C. Mignerey, A. Turgut, J. A. Schindall, and D. J.

Goldstein. Evaluation of relative entropy for distributed passive detection of weak acoustic signals. *IEEE J. Oceanic Engineering*, 42(1):219–230, 2017.

[21] Y. Mo and B. Sinopoli. Kalman filtering with intermittent observations: tail distribution and critical value. *IEEE Trans. Automat. Control*, 57(3):677–689, 2012.

[22] K. Ozcan, S. Velipasalar, and P. K. Varshney. Autonomous fall detection with wearable cameras by using relative entropy distance measure. *IEEE Trans. Human-Machine Systems*, 47(1):31–39, 2017.

[23] N. J. Ploplys, P. A. Kawka, and A. G. Alleyne. Closed-loop control over wireless networks. *Control Systems, IEEE*, 24(3):58–71, 2004.

[24] E. R. Rohr, D. Marelli, and M. Fu. Kalman filtering with intermittent observations: On the boundedness of the expected error covariance. *IEEE Trans. Automat. Control*, 59(10):2724–2738, 2014.

[25] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry. Foundations of control and estimation over lossy networks. *Proc. the IEEE*, 95(1):163–187, 2007.

[26] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry. Kalman filtering with intermittent observations. *IEEE Trans. Automat. Control*, 49(9):1453–1464, 2004.

[27] Y. Song, Z. Wang, D. Ding, and G. Wei. Robust model predictive control under redundant channel transmission with applications in networked DC motor systems. *Int. J. Robust Nonlin. Control*, 26(18):3937–3957, 2016.

[28] J. L. Speyer and W. H. Chung. *Stochastic Processes, Estimation, and Control*, volume 17. SIAM, 2008.

[29] B. Tang, M. M. Naghsh, and J. Tang. Relative entropy-based waveform design for MIMO radar detection in the presence of clutter and interference. *IEEE Trans. Signal Processing*, 63(14):3783–3796, 2015.

[30] S. Tong, Z. Li, Y. Zhu, and N. V. Thakor. Describing the nonstationarity level of neurological signals based on quantifications of time–frequency representation. *IEEE Trans. Biomedical Engineering*, 54(10):1780–1785, 2007.

[31] D. Wang, Z. Wang, B. Shen, F. E. Alsaadi, and T. Hayat. Recent advances on filtering and control for cyber-physical systems under security and resource constraints. *J. Franklin Inst.*, 353(11):2451–2466, 2016.

[32] S. Wang, H. Fang, and X. Liu. Distributed state estimation for stochastic non-linear systems with random delays and packet dropouts. *IET Control Theory Appl.*, 9(18):2657–2665, 2015.

[33] X. Wang. Recursive algorithms for linear LMSE estimators under uncertain observations. *IEEE Trans. Autom. Control*, 29(9):853–854, 1984.

[34] Z. Wang, D. W. C. Ho, Y. Liu, and X. Liu. Robust H_∞ control for a class of nonlinear discrete time-delay stochastic systems with missing measurements. *Automatica*, 45(3):684–691, 2009.

[35] K. You, M. Fu, and L. Xie. Mean square stability for Kalman filtering with Markovian packet losses. *Automatica*, 47(12):2647–2657, 2011.

[36] H. Yuan and Y. Xia. Secure filtering for stochastic non-linear systems under multiple missing measurements and deception attacks. *IET Control Theory Appl.*, 12(4):515–523, 2017.

[37] L. Zhang, Y. Zhu, and W. X. Zheng. State estimation of discrete-time switched neural networks with multiple communication channels. *IEEE Trans. Cybern.*, 47(4):1028–1040, 2017.

[38] Y. Zhu, L. Zhang, and W. X. Zheng. Distributed H_∞ filtering for a class of discrete-time Markov jump Lur'e systems with redundant channels. *IEEE Trans. Ind. Electron.*, 63(3):1876–1885, 2016.