Automatica 47 (2011) 2647–2657

Contents lists available at SciVerse ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Mean square stability for Kalman filtering with Markovian packet losses[☆]

Keyou You[a], Minyue Fu[b,c], Lihua Xie[a,1]

[a] EXQUISITUS, Centre for E-City, School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore

[b] Department of Control Science and Engineering, Zhejiang University, Hangzhou, China

[c] School of Electrical Engineering and Computer Science, University of Newcastle, NSW 2308, Australia

ARTICLE INFO

Article history:
Received 15 November 2010
Received in revised form
18 April 2011
Accepted 8 June 2011
Available online 6 October 2011

Keywords:
Stochastic linear systems
Kalman filtering
Error covariance matrices
Stability
Markovian packet losses
Stopping time

ABSTRACT

This paper studies the stability of Kalman filtering over a network subject to random packet losses, which are modeled by a time-homogeneous ergodic Markov process. For second-order systems, necessary and sufficient conditions for stability of the mean estimation error covariance matrices are derived by taking into account the system structure. While for certain classes of higher-order systems, necessary and sufficient conditions are also provided to ensure stability of the mean estimation error covariance matrices. All stability criteria are expressed by simple inequalities in terms of the largest eigenvalue of the open loop matrix and transition probabilities of the Markov process. Their implications and relationships with related results in the literature are discussed.

1. Introduction

This work is a contribution to the stability analysis of Kalman filtering with random packet losses. A motivating example is given by a sensor and an estimator/controller communicating over a wireless channel for which the quality of the communication channel varies over time because of random fading and congestion. This happens in resource limited wireless sensor networks where communication between devices are power constrained and therefore limited in range and reliability.

Kalman filtering is of great importance in networked systems due to its various applications ranging from tracking and detection to control. Recently, much attention has been paid to the stability analysis of Kalman filtering with intermittent observations; see Hespanha, Naghshtabrizi, and Xu (2007); Schenato, Sinopoli, Franceschetti, Poolla, and Sastry (2007) and the references therein. The pioneering work (Sinopoli et al., 2004) studies the optimal state estimation problem for a discrete-time linear stochastic system under the assumption that the raw measurements of the system are randomly dropped. By modeling the packet loss process as an independent and identically distributed (i.i.d.) Bernoulli process, they prove the existence of a critical packet loss rate above which the mean state estimation error covariance matrices will diverge. However, they are unable to exactly quantify the critical loss rate for general systems except providing its lower and upper bounds, which are attainable under some special cases, e.g., the lower bound is tight if the observation matrix is invertible. A less restrictive condition is provided by Plarre and Bullo (2009), where invertibility on the observable subspace is required. Mo and Sinopoli (2010) explicitly characterize the loss rate for a wider class of systems, including second-order systems and the so-called non-degenerate higher-order systems. A remarkable discovery by Mo and Sinopoli (2010) is that there are counterexamples of second-order systems for which the lower bound given by Sinopoli et al. (2004) is not tight.

Since the communication channel state generally does not vary independently in time, an i.i.d. process is inadequate to describe the packet loss process. To capture the possible temporal correlation of network conditions, a time-homogeneous binary Markov process is adopted to model the packet loss process by Huang and Dey (2007). This is usually called the Gilbert–Elliott channel model. Under i.i.d. packet losses, stability of the estimation error covariance matrices in the mean sense may be effectively analyzed by a modified discrete-time Riccati recursion. In contrast, this approach is no longer feasible for the Markovian packet loss model, rendering the stability analysis more challenging. Due to

☆ This work was supported in part by the National Natural Science Foundation of China under grants NSFC 60828006 and 61120106011. The material in this paper was partially presented at the 18th IFAC World Congress, August 28–September 2, 2011, Milan, Italy. This paper was recommended for publication in revised form by Associate Editor Michael V. Basin under the direction of Editor Ian R. Petersen.

E-mail addresses: youk0001@ntu.edu.sg (K. You), minyue.fu@newcastle.edu.au (M. Fu), elhxie@ntu.edu.sg, ELHXIE@ntu.edu.sg (L. Xie).

[1] Tel.: +65 67904524; fax: +65 67920415.

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.09.015

the temporal correlation of the Markov process, the generalization of the results of Mo and Sinopoli (2010) to the Markov packet loss model is far from trivial. In Huang and Dey (2007), an interesting notion of peak covariance stability in the mean sense is introduced. They give a sufficient condition for this stability notion for vector systems, which is also necessary for systems with observation index one. A less conservative sufficient condition for the peak covariance stability under some cases is provided by Xie and Xie (2008). However, these works do not exploit the system structure and fail to offer necessary and sufficient conditions for the peak covariance stability, except for the special systems with observation index one. In addition, they are unable to characterize the relationship between the peak covariance stability and the usual stability of the estimation error covariance matrices for vector systems. Thus, the usual stability condition for the mean estimation error covariance matrices of vector systems with Markovian packet losses is yet to be known.

Note that if the sensor is allowed to equip with computing power and memory to preprocess the measurements, the effects of i.i.d. packet losses on the mean square stability of the optimal state estimate can be significantly reduced (Schenato, 2008). Under the same assumption on the sensor's capability, the stabilizability of a discrete-time linear time-invariant system using two remote sensors over packet dropping channels is analyzed by Gupta, Martins, and Baras (2009). Moreover, there are some other probabilistic descriptions to examine the behavior of the estimation error covariance matrices, which are stochastic due to random packet losses. In Shi, Epstein, and Murray (2010), they study the performance of Kalman filtering by considering a different metric $\mathbb{P}(P_k \leq M)$, i.e., the probability that the one-step prediction error covariance matrix P_k is bounded by a given positive definite matrix M, which is related to finding the cumulative distribution of P_k. This probability could be exactly computed for scalar systems and only has lower and upper bounds for vector systems (Shi et al., 2010). Another performance metric called the stochastic boundedness is introduced by Kar, Sinopoli, and Moura (2010) for an i.i.d. packet loss model. It is worth pointing out that under different metrics or scenarios, the effects of random packet losses on performance would be substantially different. Other related works include Censi (2011); Dey, Leong, and Evans (2009); Epstein, Shi, Tiwari, and Murray (2008); Gupta, Dana, Hespanha, Murray, and Hassibi (2009); Hu and Yan (2007); Kluge, Reif, and Brokate (2010); Mostofi and Murray (2009); Sun, Xie, Xiao, and Soh (2008); Trivellato and Benvenuto (2010); Xiao, Xie, and Fu (2009).

The present work continues to study the stability of Kalman filtering with Markovian packet losses in transmitting the raw measurements. Instead of directly analyzing a random Riccati recursion as in Huang and Dey (2007), the system structure is exploited to investigate the effects of Markovian packet loss on stability. Motivated by You and Xie (2011), we first investigate the stability of the estimation error covariance matrices at packet reception times and introduce the notion of *stability in stopping times*. It turns out that this problem is equivalent to the stability of Kalman filtering for a stochastically time-varying linear system, whose studies can be traced back to Bougerol (1993). However, the framework by Bougerol (1993) is quite general and not suitable to quantify the effects of Markovian packet losses on stability. Another stability notion is the usual stability of the mean state estimation error covariance matrices in the literature, which is referred to as *stability in sampling times* for comparison. Although the first stability notion (in stopping times) deals with stability of a randomly down-sampled system obtained by down-sampling the discrete-time system at the stopping times, both the aforementioned stability notions are shown to be equivalent. Thus, the mean estimation error covariance matrices for the down-sampled system and the original discrete-time system are with the same stability property. Apart from the theoretical merit on its own, this result allows us to relatively easily analyze the stability of the estimation error covariance matrices because the first stability notion is generally easier to study.

For second-order systems, necessary and sufficient conditions for the stability of the mean estimation error covariance matrices under different system structures are derived, respectively. For certain classes of higher-order systems, including that each unstable eigenvalue of the open-loop matrix associates with only one Jordan block and has a distinct magnitude and also non-degenerate systems (Mo & Sinopoli, 2010), a simple necessary and sufficient condition for the stability of the mean estimation error covariance matrices is obtained. All stability criteria of this work are described by simple strict inequalities in terms of the largest eigenvalue of the open loop matrix and transition probabilities of the Markov process, rather than an infinite sum as in Huang and Dey (2007) and Xie and Xie (2008). Thus, the effect of the Markov packet losses and the largest unstable eigenvalue on stability could be easily understood for the above systems. However, based on the existing results such as Huang and Dey (2007) and Xie and Xie (2008), it is unclear whether the stability condition relies on other unstable eigenvalues of the open-loop matrix, besides the largest one.

The rest of the paper is organized as follows. The problem under consideration is precisely formulated in Section 2, where two stability notions are introduced. In Section 3, a necessary condition for both stability notions of vector systems is derived, from which the equivalence between the two stability notions is established. Necessary and sufficient conditions for the stability of the mean estimation error covariance matrices of second-order systems is provided in Section 4. The necessary condition presented in Section 3 is proved to be sufficient for certain classes of higher-order systems in Section 5. Illustrative examples are presented in Section 6. Concluding remarks are drawn in Section 7. To improve the readability of the paper, most of proofs are moved to Appendix. A preliminary version of this paper on the scalar measurement can be found in You, Fu, and Xie (2011).

Notation: For a symmetric matrix M, $M \geq 0$ ($M > 0$) means that the matrix is positive semi-definite (definite), and the relation $M_1 \geq M_2$ for symmetric matrices means that $M_1 - M_2 \geq 0$. The sets of nonnegative integers, real numbers and complex numbers are denoted by $\mathbb{N}, \mathbb{R}$ and $\mathbb{C}$, respectively. $\mathrm{tr}(M)$ represents the summation of all the diagonal elements of M while $\det(M)$ returns its determinant. $\|\cdot\|$ and $\rho(\cdot)$ denote the spectral norm and the spectral radius of a matrix, respectively.

2. Problem formulation

Consider a discrete-time stochastic linear system:

$$\begin{cases} x_{k+1} = Ax_k + w_k; \\ y_k = Cx_k + v_k, \end{cases} \tag{1}$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ are the vector state and measurement. $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are white Gaussian noises with zero means and covariance matrices $Q > 0$ and $R > 0$, respectively. C is of full row rank, i.e., $\mathrm{rank}(C) = m \leq n$. The initial state x_0 is a random Gaussian vector of mean $\bar{x}_0$; and the covariance matrix $P_0 > 0$. Moreover, w_k, v_k and x_0 are mutually independent.

We focus on a network environment where the raw measurements of the system are transmitted to an estimator via an unreliable communication channel; see Fig. 1. Due to random fading

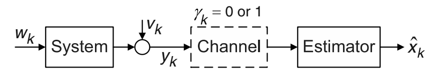

Fig. 1. Network configuration.

and/or congestion of the communication channel, packets may be lost while in transit through the channel. Different from You and Xie (2010, 2011), the present work ignores other effects such as quantization, transmission errors and data delays. The packet loss process is modeled by a time-homogeneous binary Markov process $\{\gamma_k\}_{k\geq 0}$, which is more general and realistic than the i.i.d. case studied by Sinopoli et al. (2004) due to possible temporal correlation of network conditions. Furthermore, assume that $\{\gamma_k\}_{k\geq 0}$ does not contain any information of the system. Let $\gamma_k = 1$ indicate that the packet containing the information of y_k has been successfully delivered to the estimator while $\gamma_k = 0$ corresponds to the loss of the packet. In addition, the Markov process has a transition probability matrix given by

$$\Pi^+ = (\mathbb{P}\{\gamma_{k+1} = j|\gamma_k = i\})_{i,j\in\mathbb{S}} = \begin{bmatrix} 1-q & q \\ p & 1-p \end{bmatrix}, \quad (2)$$

where $\mathbb{S} \triangleq \{0, 1\}$ is the state space of the Markov process. To avoid any trivial case, the *failure rate* p and *recovery rate* q are assumed to be strictly positive and less than 1, i.e., $0 < p, q < 1$, so that the Markov process $\{\gamma_k\}_{k\geq 0}$ is ergodic. Obviously, a smaller value of p and a larger value of q indicate a more reliable communication link.

Denote the common probability space by $(\Sigma, \mathcal{F}, \mathbb{O})$ for all random variables in the paper, where Ω is the space of elementary events, $\mathcal{F}$ is the underlying σ-field on Ω, and $\mathbb{P}$ is a probability measure on $\mathcal{F}$. Let $\mathcal{F}_k \triangleq \sigma(y_i\gamma_i, \gamma_i, i \leq k) \subset \mathcal{F}$ be an increasing sequence of σ-fields generated by the information received by the estimator up to time k, i.e., all events that are generated by the random variables $\{y_i\gamma_i, \gamma_i, i \leq k\}$. In the sequel, the terminology of almost everywhere (abbreviated as *a.e.*) is always with respect to (w.r.t.) the probability measure $\mathbb{P}$. Associated with the Markov process $\{\gamma_k\}_{k\geq 0}$, define a sequence of stopping times $\{t_k\}_{k\geq 0}$ adapted to the Markov process $\{\gamma_k\}_{k\geq 0}$ as follows:

$$\begin{aligned} t_0 &= 0, \\ t_1 &= \inf\{k|k \geq 1, \gamma_k = 1\}, \\ t_2 &= \inf\{k|k > t_1, \gamma_k = 1\}, \\ &\vdots \quad \vdots \\ t_k &= \inf\{k|k > t_{k-1}, \gamma_k = 1\}. \end{aligned} \quad (3)$$

By the ergodic property of the Markov process $\{\gamma_k\}_{k\geq 0}$, t_k is finite *a.e.* for any fixed k (Meyn, Tweedie, & Hibey, 1996). Thus, the integer valued sojourn times $\{\tau_k\}_{k>0}$ to denote the time duration between two successive packet received times are well-defined *a.e.*, where

$$\tau_k \triangleq t_k - t_{k-1} > 0. \quad (4)$$

With regard to the probability distribution of sojourn times $\{\tau_k\}_{k>0}$, we recall the following interesting result.

Lemma 1 (*Xie and Xie (2009)*)**.** *Conditioned on the event that $\{\gamma_0 = 1\}$, the sojourn times $\{\tau_k\}_{k>0}$ are independent and identically distributed. Furthermore, the conditional distribution of τ_1 is explicitly expressed as*

$$\mathbb{P}\{\tau_1 = i|\gamma_0 = 1\} = \begin{cases} 1-p, & i = 1; \\ pq(1-q)^{i-2}, & i > 1. \end{cases} \quad (5)$$

As in You and Xie (2011), we shall use the above lemma to establish our results. To this purpose, denote the state estimate and one-step prediction corresponding to the minimum mean square error estimator by $\hat{x}_{k|k} = \mathbb{E}[x_k|\mathcal{F}_k]$ and $\hat{x}_{k+1|k} = \mathbb{E}[x_{k+1}|\mathcal{F}_k]$, respectively. The associated estimation error covariance matrices are defined by $P_{k|k} = \mathbb{E}[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^H|\mathcal{F}_k]$ and $P_{k+1|k} = \mathbb{E}[(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^H|\mathcal{F}_k]$, where A^H is the conjugate transpose of A. From Sinopoli et al. (2004), it is known that the Kalman filter is still optimal. That is, the following recursions are in force:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \gamma_k K_k(y_k - C\hat{x}_{k|k-1}); \quad (6)$$

$$P_{k|k} = P_{k|k-1} - \gamma_k K_k C P_{k|k-1}, \quad (7)$$

where $K_k = P_{k|k-1}C^H(CP_{k|k-1}C^H + R)^{-1}$. In addition, the time update equations continue to hold: $\hat{x}_{k+1|k} = A\hat{x}_{k|k}$, $P_{k+1|k} = AP_{k|k}A^H + Q$ and $\hat{x}_{0|-1} = \bar{x}_0$, $P_{0|-1} = P_0$. For simplicity of exposition, let $P_{k+1} = P_{k+1|k}$ and $M_k = P_{t_k+1}$. To analyze the behavior of the estimation error covariance matrices, we introduce two types of stability.

Definition 1. We say that the mean state estimation error covariance matrices are stable in sampling times if $\sup_{k\in\mathbb{N}} \mathbb{E}[P_k] < \infty$[2] while they are stable in stopping times if $\sup_{k\in\mathbb{N}} \mathbb{E}[M_k] < \infty$ for any $P_0 > 0$, where the expectation is taken w.r.t. packet loss process $\{\gamma_k\}_{k\geq 0}$ with γ_0 being any Bernoulli random variable.

Here $\mathbb{E}[P_k]$ represents the mean of one-step prediction error covariance at the sampling time whereas $\mathbb{E}[M_k]$ denotes the mean of one-step prediction error covariance at the stopping time. To some extent, the former is time-driven while the latter is event-driven. Although the two stability notions have different meanings, they will be shown to be equivalent in Section 3. Our objective of this paper is to establish the equivalence between the two stability notions and derive necessary and sufficient conditions for stability. For scalar systems, the stability in sampling times has been discussed by Huang and Dey (2007) by analyzing a random Riccati recursion. Their approach is quite conservative for vector systems as they leave the system structure unexplored. In this paper, a completely different method is developed to establish the main results.

In consideration of Theorems 3 and 8 of Mo and Sinopoli (2010), there is no loss of generality to assume that:

A1: P_0, Q, R are all identity matrices with compatible dimensions.
A2: All the eigenvalues of A lie outside the unit circle.
A3: (C, A) is observable.

3. Equivalence of the two stability notions

It is generically difficult to directly study the notion of stability in sampling times (Huang & Dey, 2007). However, the two stability notions will be shown to be equivalent in this section.

For any $i \in \mathbb{S}$, denote by $\mathbb{E}^i[\cdot]$ the mathematical expectation operator conditioned on the event that $\{\gamma_0 = i\}$.

Lemma 2. *The following statements hold:*

(a) $\sup_{k\in\mathbb{N}} \mathbb{E}[P_k] < \infty$ *if and only if* $\sup_{k\in\mathbb{N}} \mathbb{E}^1[P_k] < \infty$ *and* $\sup_{k\in\mathbb{N}} \mathbb{E}^0[P_k] < \infty$.
(b) $\sup_{k\in\mathbb{N}} \mathbb{E}[M_k] < \infty$ *if and only if* $\sup_{k\in\mathbb{N}} \mathbb{E}^1[M_k] < \infty$ *and* $\sup_{k\in\mathbb{N}} \mathbb{E}^0[M_k] < \infty$.

Proof. (a) "⇐:" It is obvious since $\mathbb{E}[P_k] \leq \mathbb{E}^1[P_k] + \mathbb{E}^0[P_k]$. "⇒:" Let $\mathbb{P}\{\gamma_0 = 1\} = \mathbb{P}\{\gamma_0 = 0\} = 1/2$. Note that $P_k \geq 0$; then $\mathbb{E}[P_k] \geq \mathbb{E}^1[P_k]/2$ and $\mathbb{E}[P_k] \geq \mathbb{E}^0[P_k]/2$. (b) Similar to (a). □

Theorem 3. *Consider system (1) satisfying A1–A3 and the packet loss process of the measurements governed by a time-homogeneous Markov process with transition probability matrix (2). Then, a necessary condition for $\sup_{k\in\mathbb{N}} \mathbb{E}[M_k] < \infty$ is that $\rho(A)^2(1-q) < 1$.*

[2] This notation means that there is a positive definite $\bar{P}$ such that $\mathbb{E}[P_k] < \bar{P}$ for all $k \in \mathbb{N}$. A similar meaning applies to the notation $\sup_{k\in\mathbb{N}} \mathbb{E}[M_k] < \infty$.

Proof. Define a linear operator $g(\cdot)$ by $g(P) = APA^H + Q$ and the composite function $g \circ g(\cdot)$ by $g \circ g(P) = g(g(P)) = g^2(P)$. A similar definition applies to the notation $g^k(\cdot)$ for all $k \geq 1$. Since t_k is a stopping time, $\mathcal{F}_{t_k} \triangleq \sigma(y_i\gamma_i, \gamma_i, i \leq t_k)$ is a well defined σ-field. Noting that $P_k \geq Q = I$ for all $k \in \mathbb{N}$ and $M_k = P_{t_k+1}$, it immediately follows from the property of conditional expectation that

$$\begin{aligned}\mathbb{E}[M_{k+1}] &= \mathbb{E}\left[\mathbb{E}[M_{k+1}|\mathcal{F}_{t_k}]\right] \\ &= \mathbb{E}\left[g^{\tau_{k+1}}(M_k)\right] \geq \mathbb{E}[g^{\tau_{k+1}}(Q)] \\ &= \mathbb{E}\left[\sum_{j=0}^{\tau_{k+1}} A^j(A^j)^H\right],\end{aligned} \tag{8}$$

where the first inequality is due to that $g(\cdot)$ is a monotonically increasing function. Let $J = \mathrm{diag}(J_1, \ldots, J_d) \in \mathbb{C}^{n\times n}$ be the Jordan canonical form of A, where $J_i \in \mathbb{C}^{n_i\times n_i}$ corresponds to the eigenvalue λ_i. That is, there exists a nonsingular matrix $U \in \mathbb{R}^{n\times n}$ such that $A = UJU^{-1}$. Then, it follows that

$$\begin{aligned}\sum_{j=0}^{\tau_{k+1}} A^j(A^j)^H &= U\sum_{j=0}^{\tau_{k+1}} J^jU^{-1}U^{-H}(J^j)^HU^H \\ &\geq \lambda_{\min}(U^{-1}U^{-H})U\sum_{j=0}^{\tau_{k+1}} J^j(J^j)^HU^H,\end{aligned} \tag{9}$$

where $\lambda_{\min}(U^{-1}U^{-H}) > 0$ is the smallest eigenvalue of $U^{-1}U^{-H}$. In view of (8) and (9) and Lemma 2, it is clear that $\sup_{k\in\mathbb{N}} \mathbb{E}[M_{k+1}] < \infty$ implies that

$$\sup_{k\in\mathbb{N}} \mathbb{E}^1\left[\sum_{j=0}^{\tau_{k+1}} J^j(J^j)^H\right] < \infty. \tag{10}$$

Note that the (n_i, n_i)-th element of $\mathbb{E}^1[\sum_{j=0}^{\tau_{k+1}} J_i^j(J_i^j)^H]$ is computed by $\mathbb{E}^1[\sum_{j=0}^{\tau_{k+1}} |\lambda_i|^{2j}] = \frac{|\lambda_i^2|\mathbb{E}^1[|\lambda_i|^{2\tau_1}]-1}{|\lambda_i|^2-1}$. By (10) and the equivalence property of norms on a finite-dimensional vector space, it follows that $\frac{|\lambda_i^2|\mathbb{E}^1[|\lambda_i|^{2\tau_1}]-1}{|\lambda_i|^2-1} < \infty$. Together with Lemma 1, we have that $|\lambda_i|^2(1-q) < 1$. Since λ_i is an arbitrary eigenvalue of A, this completes the proof. □

Theorem 4. *Consider system (1) satisfying A1–A3 and the packet loss process of the measurements governed by a time-homogeneous Markov process with transition probability matrix (2). Then, a necessary condition for* $\sup_{k\in\mathbb{N}} \mathbb{E}[P_k] < \infty$ *is that* $\rho(A)^2(1-q) < 1$.

Proof. Since $P_k - P_kC^H(CP_kC^H + R)^{-1}CP_k \geq 0$ (see Xie and Xie (2008)), we obtain that for any $k > 3$,

$$\begin{aligned}P_{k+1} &\geq (1-\gamma_k)AP_kA^H + Q \\ &\geq \sum_{j=1}^{k}\left(\prod_{i=j}^{k}(1-\gamma_i)\right)A^{k-j}(A^{k-j})^H,\end{aligned} \tag{11}$$

where the second inequality is due to that $Q = I$ by A1. Denote $\pi_j^i = \mathbb{P}\{\gamma_j = i\}, i \in \{0, 1\}$ and $\pi_j = [\pi_j^0, \pi_j^1]$. By (2), we have that $\pi_{j+1} = \pi_j\Pi^+$ for any $j \in \mathbb{N}$. Together with $0 < p, q < 1$, one can test that for any finite $j > 1$, $\pi_j^i > 0$ for all $i \in \{0, 1\}$. In addition, the Markov process $\{\gamma_k\}_{k\in\mathbb{N}}$ has a unique stationary distribution $[\pi^0, \pi^1]$, i.e., $\lim_{j\to\infty} \pi_j^i = \pi^i, i \in \mathbb{S}$. By (2), we further obtain that $\pi^0 = \frac{p}{p+q} > 0$. Then, it follows that $\underline{\pi}^0 \triangleq \inf_{j\geq 1} \pi_j^0 > 0$, which further implies that for all $j \geq 2$,

$$\begin{aligned}\mathbb{E}\left[\prod_{i=j}^{k}(1-\gamma_i)\right] &\geq \mathbb{E}\left[\prod_{i=j}^{k}(1-\gamma_i)|\gamma_{j-1} = 0\right]\mathbb{P}(\gamma_{j-1} = 0) \\ &\geq \underline{\pi}^0(1-q)^{k-j}.\end{aligned} \tag{12}$$

In view of (11), we obtain that $\mathbb{E}[P_{k+1}] \geq \underline{\pi}^0\sum_{j=0}^{k-2}(1-q)^jA^j(A^j)^H$. By following a similar line of the proof in Theorem 3, it immediately yields that $\rho(A)^2(1-q) < 1$. □

Remark 5. Let $\bar{q} = \max\{q, 1-p\}$, Xie and Xie (2008) provides a necessary condition, i.e., $\rho^2(A)(1-\bar{q}) < 1$ for $\sup_{k\in\mathbb{N}} \mathbb{E}[P_k] < \infty$, which is obviously weaker than Theorem 4 if $p + q < 1$.

By the above results, the equivalence between the two stability notions is established in the following result, whose proof is given in Appendix.

Theorem 6. *Consider system (1) satisfying A1–A3 and the packet loss process of the measurements governed by a time-homogeneous Markov process with transition probability matrix (2). Then, the notions of stability in stopping times and stability in sampling times are equivalent.*

Thus, there is no loss of generality for the rest of the paper to focus on the stability in stopping times.

4. Second-order systems

Consider second-order systems with the following structure:

A4: $A = \mathrm{diag}(\lambda_1, \lambda_2)$ and $\mathrm{rank}(C) = 1$, where $\lambda_2 = \lambda_1\exp(\frac{2\pi r}{d}\mathbb{i})$, $\mathbb{i}^2 = -1$, $d > r \geq 1$ and $r, d \in \mathbb{N}$ are irreducible.

Under A4, it is easy to verify that (C, A^d) is not an observable pair. This essentially indicates that the measurements received at times kd for all $k \in \mathbb{N}$ do not help to reduce the estimation error, which will become clear shortly. Thus, it is intuitive that with a smaller d, it may require a stronger condition to ensure stability of the mean estimation error covariance matrices as observability may be lost relatively easily, which is confirmed in Theorem 7.

Theorem 7. *Consider second-order system (1) satisfying A1–A3 and the packet loss process of the measurements governed by a time-homogeneous Markov process with transition probability matrix (2). Then,*

(a) *if the pair* (C, A) *satisfies A4, a necessary and sufficient condition for* $\sup_{k\in\mathbb{N}} \mathbb{E}[M_k] < \infty$ *is that* $(1 + \frac{pq}{(1-q)^2})(\rho(A)^2(1-q))^d < 1$;
(b) *otherwise, a necessary and sufficient condition for* $\sup_{k\in\mathbb{N}} \mathbb{E}[M_k] < \infty$ *is that* $\rho(A)^2(1-q) < 1$.

The proof is delivered in Appendix. By Theorem 6, the results in Theorem 7 apply to the notion of stability in sampling times as well. Some remarks are included below.

Remark 8. Since $d \geq 2$, the function $(1+\frac{pq}{(1-q)^2})(1-q)^d$ is decreasing w.r.t. $q \in (0, 1)$ but increasing w.r.t. $p \in (0, 1)$. For a communication link with a smaller p and a larger q, which corresponds to a more reliable network, a more unstable system can be tolerated without losing stability of the estimation error covariance matrices. This is consistent with our intuition.

Remark 9. If the conjugate complex eigenvalues satisfy that $\lambda_2 = \lambda_1\exp(2\pi\varphi\mathbb{i})$, where φ is an irrational number, A4 does not hold. A necessary and sufficient condition for both the types of stability is that $|\lambda_1|^2(1-q) < 1$. Under this situation, the pair (C, A^k) remains observable for all $k \geq 1$. Then, the failure rate p becomes immaterial. In Section 5, we show that even for certain classes of higher-order systems with scalar measurements, the failure rate is of little importance for stability as well.

Remark 10. In Huang and Dey (2007), they establish the equivalence of the usual stability (stability in sampling times) and the so-called peak covariance stability of the estimation error covariance matrices only for scalar systems. But for vector systems, they give a conservative sufficient condition for the peak covariance stability and do not consider the usual stability.

Remark 11. If the packet loss process is an i.i.d. process, corresponding to $q = 1 - p$ in the transition probability matrix of the Markov process, the stability criterion under A4 in Theorem 7 is reduced to that $q > 1 - \rho(A)^{-\frac{2d}{d-1}}$, which recovers the result by Mo and Sinopoli (2010). Note that under i.i.d. packet losses, a lower bound for the critical packet loss rate given by Sinopoli et al. (2004) is interpreted as $q > 1 - \rho(A)^{-2}$, which is obviously not tight for systems satisfying A4.

5. Higher-order systems

Under an i.i.d. packet loss assumption, an explicit characterization of necessary and sufficient conditions for stability of filtering error covariance for general vector linear systems is known to be extremely challenging (Mo & Sinopoli, 2010; Plarre & Bullo, 2009; Sinopoli et al., 2004). Fortunately, for certain classes of higher-order systems, where each stable eigenvalue of A^{-1} associates with only one Jordan block and has a distinct magnitude or (C, A) is a non-degenerate pair, it is possible to give a simple necessary and sufficient condition for stability of the estimation error covariance matrices. This section shows that the condition in Theorem 3 is also sufficient under certain classes of higher-order systems, whose proofs are given in Appendix. To this aim, some definitions introduced by Mo and Sinopoli (2010) are adopted.

Definition 2. The pair (C, A) is one step observable if C is of full column rank.

Definition 3. Assume that (C, A) is in diagonal standard form, i.e., $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $C = [C_1, \ldots, C_n]$. An equi-block of the system is defined as the subsystem corresponding to the block $(C_{\mathcal{I}}, A_{\mathcal{I}})$, where $\mathcal{I} = \{i_1, \ldots, i_l\} \subset \{1, \ldots, n\}$ is an index set such that $|\lambda_{i_1}| = \cdots = |\lambda_{i_l}|$ and $A_{\mathcal{I}} = \mathrm{diag}(\lambda_{i_1}, \ldots, \lambda_{i_l})$, $C_{\mathcal{I}} = [C_{i_1}, \ldots, C_{i_l}]$.

Definition 4. The system (C, A) is non-degenerate if every equi-block of the system is one step observable. Conversely, the system (C, A) is degenerate if there exists an equi-block of the system that is not one step observable.

The concept of a non-degenerate system is weaker than that of a one step observable system but stronger than an observable one.

A5: (C, A) is a non-degenerate pair.

Theorem 12. *Consider system (1) satisfying A1–A3, A5 and the packet loss process of the measurements governed by a time-homogeneous Markov process with transition probability matrix (2). Then, a necessary and sufficient condition for* $\sup_{k\in\mathbb{N}} \mathbb{E}[M_k] < \infty$ *is that* $\rho(A)^2(1-q) < 1$.

It should be noted that Theorem 7 of Mo and Sinopoli (2010) provides a necessary and sufficient condition for stability in sampling times for non-degenerate systems under i.i.d. packet losses. Their results indicate that the lower bound for the critical packet loss rate by Sinopoli et al. (2004) is tight for non-degenerate systems. While in Theorem 12, we give a necessary and sufficient condition for stability of non-degenerate systems under Markovian packet losses. Next, the necessary condition in Theorem 3 is proved be sufficient for another class of higher-order systems with the following structure.

A6: $A^{-1} = \mathrm{diag}(J_1, \ldots, J_m)$ and $\mathrm{rank}(C) = 1$, where $J_i = \lambda_i^{-1} I_i + N_i \in \mathbb{R}^{n_i \times n_i}$ and $|\lambda_i| > |\lambda_{i+1}|$. I_i is an identity matrix with a compatible dimension and the (j, k)-th element of N_i is 1 if $k = j + 1$ and 0, otherwise.

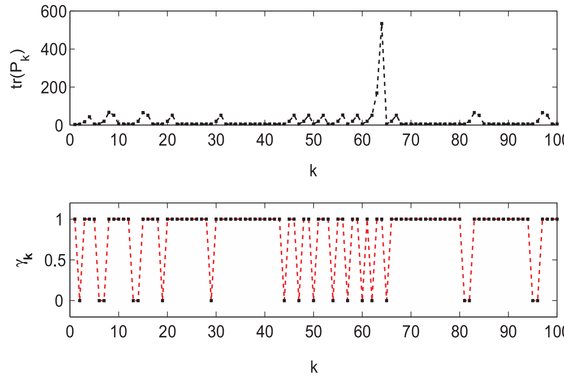

Fig. 2. A sample path with $q = 0.8$ and $p = 0.1$.

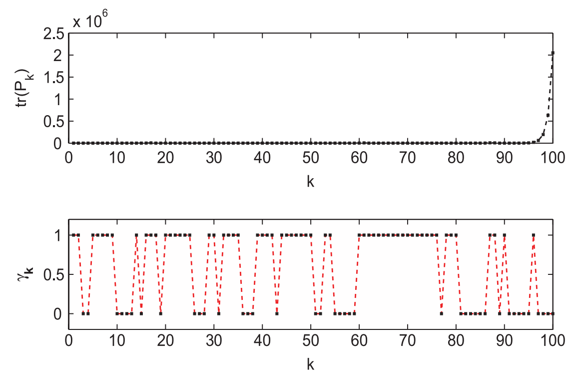

Fig. 3. A sample path with $q = 0.6$ and $p = 0.1$.

Theorem 13. *Consider system (1) satisfying A1–A3, A6 and the packet loss process of the measurements governed by a time-homogeneous Markov process with transition probability matrix (2). Then, a necessary and sufficient condition for* $\sup_{k\in\mathbb{N}} \mathbb{E}[M_k] < \infty$ *is that* $\rho(A)^2(1-q) < 1$.

Remark 14. Note that except for the case that A has n eigenvalues and each of them is with a distinct magnitude, Assumptions A5 and A6 define two disjoint classes of higher-order systems.

6. Illustrative example

Example 1. Let a second-order system be specified by

$$A = \begin{bmatrix} 1.5 & 0 \\ 0 & -1.5 \end{bmatrix} \quad \text{and} \quad C = [1 \quad 1]. \tag{13}$$

In order to achieve stability, the *failure rate p* and *recovery rate q* should satisfy that $\left(1 + \frac{pq}{(1-q)^2}\right)(1-q)^2 < 1.5^{-4} = 0.198$ by Theorem 7. Two sample paths with different recovery rates are shown in Figs. 2 and 3, which illustrate that with a smaller recovery rate, the estimation error covariance matrices have more chances to reach a high level, even diverge. Actually, it can be verified that with $q = 0.6$ and $p = 0.1$, the inequality in Theorem 7 is violated.

Example 2. The results on higher-order systems in Section 5 are applied to target tracking over a packet loss network. The dynamic of target is expressed by Singer (1970)

$$x_{k+1} = \begin{bmatrix} 1 & h & h^2 \\ 0 & 1 & h \\ 0 & 0 & 1 \end{bmatrix} x_k + w_k, \tag{14}$$

where h is the sampling period and x_k denotes the target state at time kh, including the target position, speed and acceleration. The

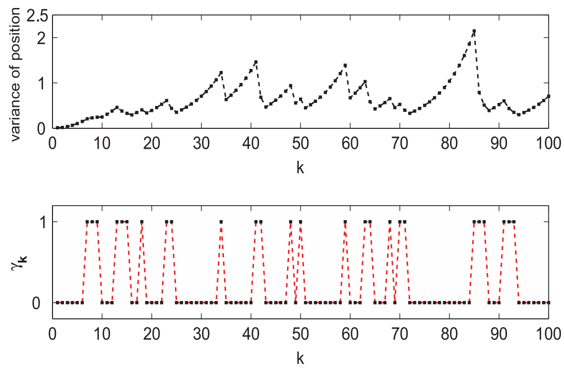

Fig. 4. A sample path with $q = 0.2$ and $p = 0.5$.

input random signal w_k is an additive white Gaussian noise. When the sampling period h is sufficiently small, the covariance of w_k is given by

$$Q = 2\alpha\sigma_m^2 \begin{bmatrix} h^5/20 & h^4/8 & h^3/6 \\ h^4/8 & h^3/3 & h^2/2 \\ h^3/6 & h^2/2 & h \end{bmatrix}, \tag{15}$$

where σ_m^2 is the variance of the target acceleration and α is the reciprocal of the maneuver time constant. The sensor periodically measures the target position with the following output equation:

$$y_k = [1 \quad 0 \quad 0]x_k + v_k, \tag{16}$$

where the measurement noise v_k is an additive white noise with variance R and independent of w_k. The initial state x_0 is a Gaussian random vector with zero mean and covariance as follows (Singer, 1970):

$$P_0 = \begin{bmatrix} R & R/h & 0 \\ R/h & 2R/h^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In this example, set $h = 0.1s, \alpha = 0.1, \sigma_m^2 = 1$ and $R = 0.01$. Although here A is marginally unstable, a scaling on A can be made as in Theorem 8 of Mo and Sinopoli (2010). Jointly with Theorem 13, it follows that $q > 0$ is sufficient to guarantee the stability of the estimation error covariance matrices. Let $q = 0.2$ and $p = 0.5$, one sample path for the tracking error variance of position is shown in Fig. 4, which illustrates that the tracking task is fulfilled.

7. Conclusion

We have examined the stability of Kalman filtering with Markovian packet losses. To analyze the random estimation error covariance matrices, two stability notions have been introduced and shown to be equivalent, which makes it relatively easier to analyze the stability of the estimation error covariance matrices. For second-order systems, necessary and sufficient conditions were obtained for ensuring stability with respect to different system structures. For certain classes of higher-order systems, a necessary and sufficient condition has been derived to guarantee the stability of estimation error covariance matrices. All results can recover the related results in the existing literature. Our future work is to find the stability conditions for general vector systems.

Appendix

Since the Markov process is temporally correlated, the proof would be more challenging than the case with i.i.d. packet losses. Before proceeding further, we need some technical lemmas.

Lemma 15 (Solo (1991)). *For any $A \in \mathbb{R}^{n\times n}$ and $\epsilon > 0$, it holds that*

$$\|A^k\| \le N\eta^k, \quad \forall k \ge 0, \tag{17}$$

where $N = \sqrt{n}(1 + \frac{2}{\epsilon})^{n-1}$ and $\eta = \rho(A) + \epsilon\|A\|$.

If A is invertible, define $\phi(k, i) = A^{t_i - t_k}$ if $k > i$ and $\phi(k, i) = I$ if $k \le i$. Let

$$\Theta_k = \sum_{i=0}^{k} \gamma_i (A^{i-k})^H C^H C A^{i-k} + (A^{-k})^H A^{-k}, \tag{18}$$

$$\Lambda_k = \sum_{j=0}^{k} \phi^H(k, j) C^H C \phi(k, j) + \phi^H(k, 0)\phi(k, 0), \tag{19}$$

$$\Xi_k = \sum_{j=0}^{k} \phi^H(j, 0) C^H C \phi(j, 0) + \phi^H(k, 0)\phi(k, 0), \tag{20}$$

$$\Xi = \sum_{j=0}^{\infty} \phi^H(j, 0) C^H C \phi(j, 0). \tag{21}$$

Lemma 16. *Under A1–A3, there exist strictly positive constant numbers α and β such that for any $k \in \mathbb{N}$,*

$$\alpha A \Lambda_k^{-1} A^H \le M_k \le \beta A \Lambda_k^{-1} A^H. \tag{22}$$

Proof. By revising Lemma 2 in Mo and Sinopoli (2010) and the fact that $\gamma_j = 0$ if $j \notin \{t_k, k \in \mathbb{N}\}$, the proof can be readily established and the details are omitted. □

By (2), it is easy to check that Ξ is invertible *a.e.* Thus, except on a set with zero probability, the inverse of Ξ is well defined. On this exceptional set, we can set Ξ^{-1} to be any value, e.g., zero matrix, as its value on a zero probability set does not affect the expectation of $\mathbb{E}[\Xi^{-1}]$.

Lemma 17. *Under A1–A3, there exist strictly positive constant numbers $\tilde{\alpha}$ and $\tilde{\beta}$ such that*

$$\tilde{\alpha} A \mathbb{E}^1[\Xi^{-1}] A^H \le \sup_{k\in\mathbb{N}} \mathbb{E}^1[M_k] \le \tilde{\beta} A \mathbb{E}^1[\Xi^{-1}] A^H. \tag{23}$$

Proof. By Lemma 1, it is clear that conditioned on the event $\{\gamma_0 = 1\}$, the following random vectors are with an identical distribution, e.g., $(\tau_k, \tau_k + \tau_{k-1}, \ldots, \tau_k + \cdots + \tau_1) \overset{d}{=} (\tau_1, \tau_1 + \tau_2, \ldots, \tau_1 + \cdots + \tau_k)$, where $\overset{d}{=}$ means equal in distribution on both its sides. Thus, it yields that $\mathbb{E}^1[\Lambda_k^{-1}] = \mathbb{E}^1[\Xi_k^{-1}]$ by (19) and (20). Jointly with Lemma 16, it follows that

$$\mathbb{E}^1[M_k] \le \beta A \mathbb{E}^1[\Xi_k^{-1}] A^H. \tag{24}$$

Under A2, it is possible to select a positive $\epsilon < \frac{1-\rho(A^{-1})}{\|A^{-1}\|}$ and $\eta = \rho(A^{-1}) + \epsilon\|A^{-1}\| < 1$; then it follows from Lemma 15 that for any $k \in \mathbb{N}$,

$$\sum_{j=k+1}^{\infty} \phi^H(j, k) C^H C \phi(j, k) \le \|C\|^2 \sum_{j=k+1}^{\infty} \|A^{(t_j - t_k)}\|^2 I$$

$$\le N\|C\|^2 \sum_{j=k+1}^{\infty} \eta^{2(t_k - t_j)} I \le \frac{N\|C\|^2}{1-\eta^2} I \triangleq \beta_0 I, \tag{25}$$

where the last inequality is due to that $\tau_k \ge 1$ for all $k \in \mathbb{N}$. Let $\beta_1 = \min(1, \beta_0^{-1})$ and $\tilde{\beta} = \beta\beta_1$; we further obtain that $\Xi_k \ge \sum_{j=0}^{k} \phi^H(j, 0) C^H C \phi(j, 0) + \beta_0^{-1} \phi^H(k, 0) (\sum_{j=k+1}^{\infty} \phi^H(j, k) C^H C \phi(j, k)) \phi(k, 0) \ge \beta_1 \Xi$, where the second inequality is due to (25). Then, the right hand side of the inequality of (23) trivially follows from (24). Similar to (24), the left hand side of (23) can be shown by using Fatou's Lemma (Ash & Doléans-Dade, 2000). □

A.1. Proof of Theorem 6

Proof. On one hand, assume that $\sup_{k\in\mathbb{N}}\mathbb{E}[P_k] < \infty$. By (2), the Markov process has a unique stationary distribution given as follows,

$$\mathbb{P}\{\gamma_\infty = i\} = \lim_{k\to\infty}\mathbb{P}\{\gamma_k = i\} = \frac{p^{1-i}q^i}{p+q}, \quad \forall i \in \mathbb{S}. \tag{26}$$

Consider a special case that the Markov process starts at its stationary distribution, i.e., $\mathbb{P}\{\gamma_0 = i\} = \frac{p^{1-i}q^i}{p+q}$ for all $i\in\mathbb{S}$. Then, the distribution of γ_k is the same as that of γ_0. Under this case, it can be verified that

$$\Pi^- = (\mathbb{P}\{\gamma_k = j|\gamma_{k+1} = i\})_{i,j\in\mathbb{S}} = \begin{bmatrix}1-q & q\\ p & 1-p\end{bmatrix}. \tag{27}$$

Given a measurable function $f : \mathbb{R}^{k+1} \to \mathbb{R}^{n\times n}$, we obtain that

$$\begin{aligned}
&\mathbb{E}[f(\gamma_k,\ldots,\gamma_0)]\\
&= \sum_{i_j\in\mathbb{S},0\le j\le k} f(i_k,\ldots,i_0)\mathbb{P}\{\gamma_k = i_k,\ldots,\gamma_0 = i_0\}\\
&= \sum_{i_j\in\mathbb{S},0\le j\le k} f(i_k,\ldots,i_0)\mathbb{P}\{\gamma_0 = i_0\}\\
&\quad\times\prod_{j=0}^{k-1}\mathbb{P}\{\gamma_{j+1} = i_{j+1}|\gamma_j = i_j\} \qquad (28)\\
&= \sum_{i_j\in\mathbb{S},0\le j\le k} f(i_k,\ldots,i_0)\mathbb{P}\{\gamma_k = i_0\}\\
&\quad\times\prod_{j=0}^{k-1}\mathbb{P}\{\gamma_j = i_{j+1}|\gamma_{j+1} = i_j\} \qquad (29)\\
&= \mathbb{E}[f(\gamma_0,\ldots,\gamma_k)] = \mathbb{E}[f(\gamma_1,\ldots,\gamma_{k+1})], \qquad (30)
\end{aligned}$$

where (28) follows from the Markov property of $\{\gamma_k\}_{k\ge0}$ while (29) is due to (2), (27) and that the distribution of γ_k is the same as that of γ_0. The last equality is due to the strict stationarity of the Markov process starting from its stationary distribution. By Lemma 3 of Mo and Sinopoli (2010), there exists a positive constant α_1 such that $P_{k+1} \ge \alpha_1(\sum_{i=1}^{k+1}\gamma_{k+1-i}(A^{-i})^H C^H CA^{-i} + (A^{-k-1})^H A^{-k-1})^{-1}$. Together with (30), we have that

$$\begin{aligned}
\mathbb{E}[P_{k+1}] &\ge \alpha_1\mathbb{E}\left(\sum_{i=1}^{k+1}\gamma_i(A^{-i})^H C^H CA^{-i} + (A^{-k-1})^H A^{-k-1}\right)^{-1}\\
&\ge \alpha_1\mathbb{E}\left(\sum_{i=1}^{\infty}\gamma_i(A^{-i})^H C^H CA^{-i} + (A^{-k-1})^H A^{-k-1}\right)^{-1}. \qquad (31)
\end{aligned}$$

Under A2, the term in (31) is decreasing w.r.t. k, which, jointly with monotone convergence theorem (Ash & Doléans-Dade, 2000), implies that $\sup_{k\in\mathbb{N}}\mathbb{E}[P_k] \ge \alpha_1\mathbb{E}[(\sum_{i=1}^{\infty}\gamma_i(A^{-i})^H C^H CA^{-i})^{-1}] \ge \alpha_1\mathbb{E}[\Xi^{-1}]$, where the last equality follows from the definition of Ξ in (21). Define a stopping time μ as the time at which the first packet is received, i.e.,

$$\mu = \inf\{k|\gamma_k = 1, \forall k\in\mathbb{N}\}.$$

Since μ is a stopping time adapted to the Markov process $\{\gamma_k\}_{k\ge0}$, we know that $\mathcal{G}_\mu \triangleq \sigma(\gamma_0,\ldots,\gamma_\mu)$ is a well defined σ-field. Furthermore, it follows from the property of conditional expectation that

$$\begin{aligned}
\mathbb{E}[\Xi^{-1}] &= \mathbb{E}\left[A^\mu\left(\sum_{j=0}^{\infty}\gamma_{j+\mu}(A^{-j})^H C^H CA^{-j}\right)^{-1}(A^\mu)^H\right]\\
&= \mathbb{E}\left[A^\mu\mathbb{E}\left[\left(\sum_{j=0}^{\infty}\gamma_{j+\mu}(A^{-j})^H C^H CA^{-j}\right)^{-1}|\mathcal{G}_\mu\right](A^\mu)^H\right].
\end{aligned}$$

By (2), it is clear that γ_k is a strong Markov process (Meyn et al., 1996). This implies that

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{j=0}^{\infty}\gamma_{j+\mu}(A^{-j})^H C^H CA^{-j}\right)^{-1}|\mathcal{G}_\mu\right]\\
&\quad= \mathbb{E}\left[\left(\sum_{j=0}^{\infty}\gamma_{j+\mu}(A^{-j})^H C^H CA^{-j}\right)^{-1}|\gamma_\mu\right]. \qquad (32)
\end{aligned}$$

By the definition of μ, it yields that $\gamma_\mu = 1$. Again, by the strong Markov property, it follows that the transition probability matrix of $\{\gamma_{k+\mu}\}_{k\ge0}$ is the same as that of the original Markov process $\{\gamma_k\}_{k\ge0}$. Combining the above, we obtain that $\sup_{k\in\mathbb{N}}\mathbb{E}[P_k] < \infty$ implies $\mathbb{E}^1[\Xi^{-1}] = \mathbb{E}[(\sum_{j=0}^{\infty}\gamma_{j+\mu}(A^{-j})^H C^H CA^{-j})^{-1}|\gamma_\mu = 1] < \infty$. By Lemma 17, it follows that $\sup_{k\in\mathbb{N}}\mathbb{E}^1[M_k] < \infty$. In view of Theorem 4, it implies that $\rho(A)^2(1-q) < 1$. This implies that $\mathbb{E}^0[A^\mu(A^\mu)^H] = q\sum_{i=1}^{\infty}A^i(A^i)^H(1-q)^{i-1} < \infty$. Together with $\sup_{k\in\mathbb{N}}\mathbb{E}^1[M_k] < \infty$, it can be easily established that $\sup_{k\in\mathbb{N}}\mathbb{E}^0[M_k] < \infty$. By Lemma 2, we finally obtain that $\sup_{k\in\mathbb{N}}\mathbb{E}[M_k] < \infty$.

On the other hand, assume that $\sup_{k\in\mathbb{N}}\mathbb{E}[M_k] < \infty$. By Lemmas 2 and 17, we obtain that $\mathbb{E}^1[\Xi^{-1}] < \infty$. By Theorem 3, it follows that $\rho(A)^2(1-q) < 1$. Then, one can easily show that $\mathbb{E}^0[\Xi^{-1}] < \infty$. As in the first part, consider the special case that the Markov process $\{\gamma_k\}_{k\ge0}$ starts at its stationary distribution. By Lemma 3 of Mo and Sinopoli (2010), there exists a positive constant β_2 such that $P_{k+1} \le \beta_2(\sum_{i=1}^{k+1}\gamma_{k+1-i}(A^{-i})^H C^H CA^{-i} + (A^{-k-1})^H A^{-k-1})^{-1}$. Together with (30), we have that

$$\begin{aligned}
\mathbb{E}[P_{k+1}] &\le \beta_2\mathbb{E}\left(\sum_{i=1}^{k+1}\gamma_i(A^{-i})^H C^H CA^{-i} + (A^{-k-1})^H A^{-k-1}\right)^{-1}\\
&= \beta_2 A\mathbb{E}[\Theta_k^{-1}]A^H, \qquad (33)
\end{aligned}$$

where the last equality is due to the strict stationarity of the Markov process as it starts from its stationary distribution and Θ_k is defined (18).

Similar to (25), there exists a positive number β_3 such that $\sum_{j=1}^{\infty}\gamma_{k+j}(A^{-j})^H C^H CA^{-j} \le \beta_3 I$. Let $\beta_4 = \min(1,\beta_3^{-1})$; we obtain that $\Theta_k \ge \beta_4\Xi$. By (33), it follows that $\mathbb{E}[P_{k+1}] \le \beta_2\beta_4^{-1}A\mathbb{E}[\Xi^{-1}]A^H < \beta_2\beta_4^{-1}A(\mathbb{E}^0[\Xi^{-1}] + \mathbb{E}^1[\Xi^{-1}])A^H < \infty$ for all $k\in\mathbb{N}$. Note that here $\mathbb{E}[P_k]$ is taken w.r.t. the Markov process $\{\gamma_k\}_{k\ge0}$ with the distribution of γ_0 being the stationary distribution. Jointly with (26), we obtain that $\mathbb{E}^0[P_k] < \infty$ and $\mathbb{E}^1[P_k] < \infty$ for all $k\in\mathbb{N}$. By Lemma 2, the proof is completed. □

A.2. Proof of Theorem 7

Proof. Define the integer valued set $\mathcal{S}_d = \{kd|\forall k \in \mathbb{N}\}$ and $\theta = \sum_{j\in\mathcal{S}_d}\mathbb{P}\{\tau_1 = j|\gamma_0 = 1\}$. Let $E_k, k\ge1$ be a sequence of events defined as follows: $E_1 = \{\tau_1 \notin \mathcal{S}_d\}$, $E_k \triangleq \{\tau_1\in\mathcal{S}_d,\ldots,\tau_{k-1}\in\mathcal{S}_d,\tau_k\notin\mathcal{S}_d\}$, for all $k\ge2$. By Lemma 1, it is obvious that $\mathbb{P}(E_k|\gamma_0 = 1) = \theta^{k-1}(1-\theta)$ and $E_i\bigcap E_j = \emptyset$ if $i\ne j$. Let $F_k = \bigcup_{j=1}^{k}E_j$

and $F = \bigcup_{j=1}^{\infty} E_j$; it follows that F_k asymptotically increases to F and $\mathbb{P}(F|\gamma_0 = 1) = \mathbb{P}(\bigcup_{j=1}^{\infty} E_j|\gamma_0 = 1) = \sum_{j=1}^{\infty}\mathbb{P}(E_j|\gamma_0 = 1) = 1$. Define the indicator function $1_{F_k}(w)$ which is one if $w \in F_k$, otherwise 0. It is clear that $1_{F_k} = \sum_{j=1}^{k} 1_{E_j}$ asymptotically increases to 1_F. Since $\mathbb{P}(F|\gamma_0 = 1) = 1$, then $1_F = 1$ *a.e.* on $\{\gamma_0 = 1\}$. Together with the monotone convergence theorem (Ash & Doléans-Dade, 2000), it follows that $\mathbb{E}^1[\Xi^{-1}] = \mathbb{E}^1[\Xi^{-1}1_F] = \mathbb{E}^1[\Xi^{-1}(\lim_{k\to\infty} 1_{F_k})] = \lim_{k\to\infty}\sum_{j=1}^{k}\mathbb{E}^1[\Xi^{-1}1_{E_j}]$.

Proof of part (a). "$\Leftarrow$:" By (21), it is clear that

$$\mathbb{E}^1[\Xi^{-1}1_{E_j}] \le \mathbb{E}^1\left[\left(\sum_{i=j-1}^{j}\phi^H(i,0)C^H C\phi(i,0)\right)^{-1} 1_{E_j}\right].$$

Define $C = [c_1, c_2]$, we can compute that

$$\sum_{i=j-1}^{j}\phi^H(i,0)C^H C\phi(i,0) = \phi^H(j-1,0)\begin{bmatrix} c_1 & \\ & c_2\end{bmatrix} \times \begin{bmatrix} 1+\lambda_1^{-2\tau_j} & 1+\lambda_1^{-\tau_j}\lambda_2^{-\tau_j} \\ 1+\lambda_1^{-\tau_j}\lambda_2^{-\tau_j} & 1+\lambda_2^{-2\tau_j}\end{bmatrix}\begin{bmatrix} c_1 & \\ & c_2\end{bmatrix}\phi(j-1,0). \quad (34)$$

Define $\Sigma_j = \begin{bmatrix} 1+\lambda_1^{-2\tau_j} & 1+\lambda_1^{-\tau_j}\lambda_2^{-\tau_j} \\ 1+\lambda_1^{-\tau_j}\lambda_2^{-\tau_j} & 1+\lambda_2^{-2\tau_j}\end{bmatrix}$; then if $\tau_j \notin \mathcal{S}_d$, it yields that $\Sigma_j^{-1} \le \frac{4}{\lambda_1^{-2\tau_j}+\lambda_2^{-2\tau_j}-2\lambda_1^{-\tau_j}\lambda_2^{-\tau_j}} I \le \frac{2|\lambda_1|^{2\tau_j}}{1-\cos(\frac{2\pi}{d})} I$. Let $c = \max(c_1^{-2}, c_2^{-2})$; it follows from (34) that if $\tau_j \notin \mathcal{S}_d$, then $(\sum_{i=j-1}^{j}\phi^H(i,0)C^H C\phi(i,0))^{-1} \le \frac{2c|\lambda_1|^{2\tau_j}}{1-\cos(\frac{2\pi}{d})} I$. Combining the above, we get that $\mathbb{E}^1[\Xi^{-1}] \le \frac{2cI}{1-\cos(\frac{2\pi}{d})}\lim_{k\to\infty}\sum_{j=1}^{k}\mathbb{E}^1[|\lambda_1|^{2\tau_j}1_{E_j}]$. By Lemma 1, the following statements are in force:

$$\begin{aligned}
&\lim_{k\to\infty}\sum_{j=1}^{k}\mathbb{E}^1[|\lambda_1|^{2\tau_j}1_{E_j}] \\
&= \lim_{k\to\infty}\sum_{j=1}^{k}\mathbb{E}^1\left[\left(\prod_{i=1}^{j-1}|\lambda_1|^{2\tau_i}1_{\{\tau_i\in\mathcal{S}_d\}}\right)|\lambda_1|^{2\tau_j}1_{\{\tau_j\notin\mathcal{S}_d\}}\right] \\
&\le \lim_{k\to\infty}\mathbb{E}^1[|\lambda_1|^{2\tau_1}]\sum_{j=1}^{k}(\mathbb{E}^1[|\lambda_1|^{2\tau_1}1_{\{\tau_1\in\mathcal{S}_d\}}])^{j-1}, \qquad (35)
\end{aligned}$$

which is finite if and only if $\mathbb{E}^1[|\lambda_1|^{2\tau_1}] < \infty$ and $\mathbb{E}^1[|\lambda_1|^{2\tau_1}1_{\{\tau_1\in\mathcal{S}_d\}}] < 1$. After some algebraic manipulations, it is easy to verify that $(1+\frac{pq}{(1-q)^2})(|\lambda_1|^2(1-q))^d < 1$ is equivalent to that $|\lambda_1|^2(1-q) < 1$ and $\frac{pq}{(1-q)^2}\frac{(|\lambda_1|^2(1-q))^d}{1-(|\lambda_1|^2(1-q))^d} < 1$. Together with Lemma 1, it implies that $\mathbb{E}^1[|\lambda_1|^{2\tau_1}] < \infty$ and

$$\mathbb{E}^1[|\lambda_1|^{2\tau_1}1_{\{\tau_1\in\mathcal{S}_d\}}] = \frac{pq}{(1-q)^2}\frac{(|\lambda_1|^2(1-q))^d}{1-(|\lambda_1|^2(1-q))^d} < 1.$$

Then, we conclude that $\mathbb{E}^1[\Xi^{-1}] < \infty$. By Lemma 17, it follows that $\sup_{k\in\mathbb{N}}\mathbb{E}^1[M_k] < \infty$. Observe that $|\lambda_1|^2(1-q) < 1$, it is easy to show that $\sup_{k\in\mathbb{N}}\mathbb{E}^0[M_k] < \infty$. By Lemma 2, we obtain that $\sup_{k\in\mathbb{N}}\mathbb{E}[M_k] < \infty$.

"$\Rightarrow$:" Denote $\Xi'_k = \sum_{j=0}^{k}\phi^H(j,0)C^H C\phi(j,0)$. In view of (25), it is easy to derive that $\Xi = \Xi'_{j-1} + \phi^H(j,0)(C^H C + \sum_{i=j+1}^{\infty}\phi^H(i,j)C^H C\phi(i,j))\phi(j,0) \le \Xi'_{j-1} + \phi^H(j,0)(C^H C + \beta_0 I)\phi(j,0)$, where β_0 is given in (25). Let $\beta_5^{-1} = \max(\frac{1}{1-|\lambda_1|^{-2}}, 1, \beta_0)$; it follows that if $1_{E_j} = 1$, then

$$\begin{aligned}
\Xi^{-1} &\ge (\Xi'_{j-1} + \phi^H(j,0)(C^H C + \beta_0 I)\phi(j,0))^{-1} \\
&= \left(\sum_{i=0}^{j-1}|\lambda_1|^{-2t_i}C^H C + \phi^H(j,0)(C^H C+\beta_0 I)\phi(j,0)\right)^{-1} \\
&\ge \left(\frac{1}{1-|\lambda_1|^{-2}}C^H C + \phi^H(j,0)(C^H C+\beta_0 I)\phi(j,0)\right)^{-1} \\
&\ge \beta_5(C^H C + \phi^H(j,0)(C^H C + I)\phi(j,0))^{-1}. \qquad (36)
\end{aligned}$$

By the definition of the indicator function, it is clear that $\Xi^{-1}1_{E_j} \ge \beta_5(C^H C + \phi^H(j,0)(C^H C + I)\phi(j,0))^{-1}1_{E_j}$. In view of Lemma 17, then $\sup_{k\in\mathbb{N}}\mathbb{E}^1[M_k] < \infty$ is equivalent to that $\mathbb{E}^1[\Xi^{-1}] < \infty$. This implies that $\lim_{k\to\infty}\sum_{j=1}^{k}\mathbb{E}^1[(C^H C + \phi^H(j,0)(C^H C + I)\phi(j,0))^{-1}1_{E_j}] < \infty$. By some manipulations, there exists a positive constant $\beta_6 > 0$ such that $\mathrm{tr}((C^H C + \phi^H(j,0)(C^H C + I)\phi(j,0))^{-1}1_{E_j}) \ge \beta_6|\lambda_1|^{2t_j}1_{E_j}$. Thus, we obtain that

$$\begin{aligned}
&\lim_{k\to\infty}\sum_{j=1}^{k}\mathbb{E}^1[|\lambda_1|^{2t_j}1_{E_j}] = \mathbb{E}^1[|\lambda_1|^2 1_{\{\tau_1\notin\mathcal{S}_d\}}] \\
&\times\lim_{k\to\infty}\sum_{j=1}^{k}(\mathbb{E}^1[|\lambda_1|^2 1_{\{\tau_1\in\mathcal{S}_d\}}])^{j-1} < \infty. \qquad (37)
\end{aligned}$$

Finally, as in the proof of sufficiency, one can easily derive that $(1+\frac{pq}{(1-q)^2})(|\lambda_1|^2(1-q))^d < 1$.

Proof of part (b). "$\Leftarrow$:" Without loss of generality, only the following cases need to be discussed.

(i) If $\mathrm{rank}(C) = 2$ or A has two eigenvalues but with distinct magnitudes, this indicates that (C, A) is a non-degenerate pair. It is proved in Theorem 12.

(ii) If $\mathrm{rank}(C) = 1$ and A contains two identical eigenvalues, it is proved in Theorem 13. Note that for this case, A cannot be of the form $A = \lambda_1 I$ for it leads to the pair (C, A) unobservable. Thus, A must contain exactly an elementary Jordan block.

(iii) If $\mathrm{rank}(C) = 1$ and $A = \mathrm{diag}(\lambda_1, \lambda_2)$, where $\lambda_2 = \lambda_1\exp(2\pi\varphi\mathfrak{i})$ and φ is an irrational number. Since φ is an irrational number and the set of rational numbers is dense, we can find a sequence of rational numbers $\{\varphi_k = \frac{r_k}{d_k}\}_{k\ge 0}$ such that $\lim_{k\to\infty}\varphi_k = \varphi$, the integers r_k and d_k are irreducible and d_k goes into infinity as $k\to\infty$. Note that $|\lambda_1|^2(1-q) < 1$, there must exist a positive integer, denoted by d_{k_0}, such that $(1+\frac{pq}{(1-q)^2})(|\lambda_1|^2(1-q))^{d_{k_0}} < 1$. Then, the rest of the proof follows similarly as the proof of sufficiency of part (a).

"$\Rightarrow$:" It directly follows from Theorem 3. □

A.3. Proofs of results in Section 5

Lemma 18 (Mo and Sinopoli (2010)). *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A and $|\lambda_1| \ge \cdots \ge |\lambda_n|$. If the pair (C, A) satisfies A2–A3 and A5, then the following inequality holds*

$$\limsup_{\Delta_1,\ldots,\Delta_n\to\infty}\frac{\left(\sum_{j=1}^{n}(A^{-k_j})^H C^H C A^{-k_j}\right)^{-1}}{\prod_{j=1}^{n}|\lambda_j|^{2\Delta_j}} \le \beta_7 I, \qquad (38)$$

where β_7 is a positive constant, $k_1 < k_2 < \cdots < k_n \in \mathbb{N}$, $\Delta_1 = k_1$, $\Delta_j = k_j - k_{j-1}$ for all $j \in \{2, \ldots, n\}$.

Proof of Theorem 12. "$\Leftarrow$:" By Lemma 18, there exists a sufficiently large $\Delta > 0$ such that for all $\Delta_j > \Delta$, it holds

$$\left(\sum_{j=1}^{n}(A^{-k_j})^H C^H C A^{-k_j}\right)^{-1} \leq \beta_7 \prod_{j=1}^{n}|\lambda_j|^{2\Delta_j} I.$$

Now, select $k_j = t_{i_j}$, where $i_1 > \Delta$, $i_j - i_{j-1} > \Delta$ for all $j \in \{2, \ldots, n\}$ and t_{i_j} is a stopping time defined in (4). Then, it is obvious that $t_{i_j} - t_{i_{j-1}} \geq i_j - i_{j-1} > \Delta$, which jointly with (21), implies that

$$\begin{aligned}
&\mathbb{E}^1[\Xi^{-1}] \\
&\leq \beta_7 \mathbb{E}^1\left[\prod_{j=1}^{n}|\lambda_j|^{2\Delta_j}\right] I = \beta_7 \mathbb{E}^1\left[\prod_{j=1}^{n}|\lambda_j|^{2(t_{i_j}-t_{i_{j-1}})}\right] I \\
&= \beta_7 \prod_{j=1}^{n}(\mathbb{E}^1[|\lambda_j|^{2\tau_1}])^{i_j - i_{j-1}} I < \infty,
\end{aligned} \tag{39}$$

where the last equality is due to Lemma 1 and we use the fact that $|\lambda_1|^2(1-q) < 1$ in the last inequality. By Lemma 17, it follows that $\sup_{k\in\mathbb{N}} \mathbb{E}^1[M_k] < \infty$. Together with that $|\lambda_1|^2(1-q) < 1$, it is easy to establish that $\sup_{k\in\mathbb{N}} \mathbb{E}^0[M_k] < \infty$. The rest of the proof follows from Lemma 2.

"$\Rightarrow$:" It is proved in Theorem 3. □

The proof of Theorem 13 is much more involved and depends on the following lemmas, which are devoted to establishing a similar result as (38) under A2–A3 and A6.

Lemma 19. *For any integer k_i such that $k_{i+1} > k_i$, let $\mathcal{B} \in \mathbb{R}^{n\times n}$ be a matrix with its (i,j)-th element given by $\mathcal{B}_{ij} = \binom{k_i}{j-1}$. Then, the determinant of $\mathcal{B}$ is computed as*

$$\det(\mathcal{B}) = \frac{1}{\prod_{i=0}^{n-1} i!} \prod_{1\leq j<i\leq n}(k_i - k_j),$$

where $i!$ is the factorial of a positive integer i.

Proof. It is clear that $\det(\mathcal{B})$ is an alternative, i.e., swapping the i-th and j-th rows is the same as changing values of k_i and k_j. Moreover, $\det(\mathcal{B})$ is an $(n-1)$-th order multivariate polynomial in $k_1, \ldots, k_n$. For example, $\det(\mathcal{B})$ is an $(n-1)$-th order polynomial in k_i when all $k_j, j \neq i$ are fixed. Combining these two properties, we obtain that $\det(\mathcal{B})$ contains $\prod_{1\leq j<i\leq n}(k_i - k_j)$ as a factor. Furthermore, $\prod_{1\leq j<i\leq n}(k_i - k_j)$ is the only factor of $\det(\mathcal{B})$, modulo a constant α_n, due to that $\prod_{1\leq j<i\leq n}(k_i - k_j)$ and $\det(\mathcal{B})$ are both of $(n-1)$-th order, from which we get the following equality:

$$\det(\mathcal{B}) = \alpha_n \prod_{1\leq j<i\leq n}(k_i - k_j). \tag{40}$$

It remains to show that $\alpha_n = 1/\prod_{i=0}^{n-1} i!$. We do so by mathematical induction. For $n = 1$, $\det \mathcal{B}(k_1) = 1$. The factor $\prod_{1\leq j<i\leq n}(k_i - k_j)$ is void and $1/\prod_{i=0}^{n-1} i! = 1$. Thus, $\alpha_1 = 1$, which is correct.

Given $n = m$; suppose it holds that $\alpha_m = 1/\prod_{i=0}^{m-1} i!$. Then, for $n = m+1$, let $k_1, \ldots, k_m$ be fixed and k_{m+1} go to infinity. Note that $\lim_{k_{m+1}\to\infty} \frac{\binom{k_{m+1}}{i}}{k_{m+1}^i} = \frac{1}{i!}$; we have

$$\begin{aligned}
&\lim_{k_{m+1}\to\infty} \frac{\det(\mathcal{B}(k_1,\ldots,k_{m+1}))}{k_{m+1}^m} \\
&= \lim_{k_{m+1}\to\infty}\left(\frac{\binom{k_{m+1}}{m}\det(\mathcal{B})}{k_{m+1}^m} + \sum_{i=0}^{m-1}\frac{O(k_{m+1}^i)}{k_{m+1}^m}\right) \\
&= \det(\mathcal{B}(k_1,\ldots,k_m))/m! \\
&= \frac{1}{\prod_{i=0}^{m} i!} \prod_{1\leq j<i\leq m}(k_i - k_j),
\end{aligned} \tag{41}$$

where $O(k_{m+1}^i)$ in the first equality means that $\lim_{k_{m+1}\to\infty} \frac{O(k_{m+1}^i)}{k_{m+1}^i} < \infty$. In light of (40), it yields that $\lim_{k_{m+1}\to\infty} \frac{\det(\mathcal{B}(k_1,\ldots,k_{m+1}))}{k_{m+1}^m} = \alpha_{m+1}\prod_{1\leq j<i\leq m}(k_i - k_j)$. Combining the above, we immediately obtain that $\alpha_{m+1} = 1/\prod_{i=0}^{m} i!$. Hence, $\alpha_n = 1/\prod_{i=0}^{n-1} i!$ holds for all $n \geq 1$. □

Lemma 20. *For any integer k_i such that $k_{i+1} > k_i$, let $\mathcal{B}' \in \mathbb{R}^{n\times n}$ be a matrix such that the (i,j)-th element is given by $\mathcal{B}'_{ij} = \binom{k_i}{j}$. Then, the determinant of $\mathcal{B}'$ is computed as $\det(\mathcal{B}') = (\prod_{i=1}^{n}\frac{k_i}{i!})\prod_{1\leq j<i\leq n}(k_i - k_j)$.*

Proof. Similar to the proof of Lemma 19, $\det(\mathcal{B}')$ is an alternative and n-th order multivariate polynomial. It is straightforward that $\det(\mathcal{B}')$ contains $\prod_{i=1}^{n} k_i$ as a factor. Thus, we further obtain that $\det(\mathcal{B}')$ contains $\prod_{i=1}^{n} k_i \prod_{1\leq j<i\leq n}(k_i - k_j)$ as a factor, which is also the only factor containing k_i. Hence, the following is in force: $\det(\mathcal{B}') = \alpha'_n(\prod_{i=1}^{n} k_i \prod_{1\leq j<i\leq n}(k_i - k_j))$. Using similar induction arguments as in Lemma 19, one can easily show that $\alpha'_n = 1/\prod_{i=1}^{n} i!$. □

Lemma 21. *Given any integer k_i such that $k_{i+1} > k_i > n$, let $\Delta_1 = k_1$, $\Delta_i = k_i - k_{i-1}$ if $i \geq 2$. Denote $\mathcal{O}(\{k_i\}_1^{n-1}) = [C^H, (A^{-k_1})^H C^H, \ldots, (A^{-k_{n-1}})^H C^H]^H$ and $D_\lambda(\{k_i\}_1^{n-1}) = \prod_{v=1}^{m}\lambda_v^{-k(v)+\frac{n_v(n_v-1)}{2}}$, where $k(1) = k_1 + \cdots + k_{n_1-1}$ and $k(v) = k_{n_1+\cdots+n_{v-1}} + \cdots + k_{n_1+\cdots+n_v-1}$ if $v \geq 2$. Under A2 and A6, we can asymptotically compute the determinant of $\mathcal{O}(\{k_i\}_1^{n-1})$. In particular, there exists a multivariate polynomial $\psi(\{k_i\}_1^{n-1})$ w.r.t. $\{k_i\}_1^{n-1}$ and independent of λ_i such that*

$$\lim_{\Delta_1,\ldots,\Delta_{n-1}\to\infty} \frac{\det \mathcal{O}(\{k_i\}_1^{n-1})}{D_\lambda(\{k_i\}_1^{n-1})\psi(\{k_i\}_1^{n-1})} = 1.$$

Proof. Under A6, partition the observation matrix C in conformity with the block diagonal matrix A. Let $C_i = [c_{i1}, \ldots, c_{in_i}]$; it is easy to verify that

$$C_i N_i^k = [\underbrace{0, \ldots, 0}_{k}, c_{i1}, \ldots, c_{i(n_i-k)}]$$

for any $k \leq n_i - 1$. We further obtain that for any $k > n_i$,

$$\begin{aligned}
&C_i(\lambda_i^{-1} I_i + N_i)^k \\
&= \left[\frac{c_{i1}}{\lambda_i^k}, \frac{c_{i2}}{\lambda_i^k} + \binom{k}{1}\frac{c_{i1}}{\lambda_i^{k-1}}, \ldots, \sum_{j=0}^{n_i-1}\binom{k}{j}\frac{c_{i(n_i-j)}}{\lambda_i^{k-j}}\right] \\
&\triangleq \lambda_i^{-k}\left[1, \binom{k}{1}, \ldots, \binom{k}{n_i-1}\right]\widetilde{C}_i,
\end{aligned}$$

where $\widetilde{C}_i$ is defined as

$$\widetilde{C}_i = \mathrm{diag}(1, \lambda_i, \ldots, \lambda_i^{n_i-1})\begin{bmatrix} c_{i1} & c_{i2} & \cdots & c_{in_i} \\ & c_{i1} & \cdots & c_{i(n_i-1)} \\ & & \ddots & \\ & & & c_{i1} \end{bmatrix}$$

and $\det(\widetilde{C}_i) = \lambda_i^{n_i(n_i-1)/2} c_{i_1}^{n_i}$. By using the above property, it follows that

$$\det \mathcal{O}(\{k_i\}_1^{n-1})$$
$$= \det \begin{bmatrix} C_1 & \dots & C_m \\ C_1(\lambda_1^{-1}I_1+N_1)^{k_1} & \dots & C_m(\lambda_m^{-1}I_m+N_m)^{k_1} \\ \ddots & & \ddots \\ C_1(\lambda_1^{-1}I_1+N_1)^{k_{n-1}} & \dots & C_m(\lambda_m^{-1}I_m+N_m)^{k_{n-1}} \end{bmatrix}$$
$$= \det \left[\begin{matrix} 1 & 0 & \dots & 0 \\ \binom{k_1}{0}\lambda_1^{-k_1} & \binom{k_1}{1}\lambda_1^{-k_1} & \dots & \binom{k_1}{n_1-1}\lambda_1^{-k_1} \\ & & \ddots & \\ \binom{k_{n-1}}{0}\lambda_1^{-k_{n-1}} & \binom{k_{n-1}}{1}\lambda_1^{-k_{n-1}} & \dots & \binom{k_{n-1}}{n_1-1}\lambda_1^{-k_{n-1}} \end{matrix}\right.$$
$$\left.\begin{matrix} \dots & 1 & 0 & \dots & 0 \\ \dots & \binom{k_1}{0}\lambda_m^{-k_1} & \binom{k_1}{1}\lambda_m^{-k_1} & \dots & \binom{k_1}{n_m-1}\lambda_m^{-k_1} \\ \ddots & & & \ddots & \\ \dots & \binom{k_{n-1}}{0}\lambda_m^{-k_{n-1}} & \binom{k_{n-1}}{1}\lambda_m^{-k_{n-1}} & \dots & \binom{k_{n-1}}{n_m-1}\lambda_m^{-k_{n-1}} \end{matrix}\right]$$
$$\times \det(\operatorname{diag}(\widetilde{C}_1,\dots,\widetilde{C}_m))$$
$$\triangleq (D_1+\dots+D_m)\det(\operatorname{diag}(\widetilde{C}_1,\dots,\widetilde{C}_m)), \tag{42}$$

where D_i is the determinant of the minor of the first matrix in the previous equation, obtained by eliminating the first row and the first column in the i-th block. For example, the first block consists of the first n_1 columns and the followed n_2 columns forms the second block.

Let $\sigma = [\sigma_1,\dots,\sigma_m]$ be a permutation of $\{1,\dots,n-1\}$ such that $\#\sigma_1 = n_1-1, \#\sigma_2 = n_2,\dots,\#\sigma_m = n_m$, where $\#\sigma_i$ denotes the order of the permutation σ_i.

Then, it follows from the Leibnitz formula (Horn & Johnson, 1985) for the determinant of a matrix that

$$D_1 = \sum_{\sigma} \operatorname{sgn}(\sigma) h(k_{\sigma_j}) \left(\prod_{j=1}^m \lambda_j^{-k_{\sigma_j}}\right), \tag{43}$$

where the signature of permutation σ is denoted as $sgn(\sigma)$, which is $+1$ for even permutation and -1 for odd permutations, $\lambda_j^{-k_{\sigma_j}} = \lambda_j^{-\sum_{i\in\sigma_j} k_i}$ and $h(k_{\sigma_i})$ is a polynomial function of k_i for all $i \in \sigma_j$. The summation is taken w.r.t. all permutations with order $n-1$.

Due to that $|\lambda_1| > \dots > |\lambda_m|$, it is clear that $\lambda^{\sigma} \triangleq |\prod_{j=1}^m \lambda_j^{-k_{\sigma_j}}|$ achieves the maximum when $\lambda_1^{-k_{\sigma_1}} = \lambda_1^{-\sum_{i\in\{1,\dots,n_1-1\}} k_i}$, $\lambda_2^{-k_{\sigma_2}} = \lambda_2^{-\sum_{i\in\{n_1,\dots,n_1+n_2-1\}} k_i}, \dots, \lambda_m^{-k_{\sigma m}} = \lambda_m^{-\sum_{i\in\{n_1+\dots+n_{m-1},\dots,n-1\}} k_i}$. Thus, denote the set of permutations having the above property by $\mathcal{P}_{\sigma}^*$. Given any permutation σ which does not belong to $\mathcal{P}_{\sigma}^*$, we always have $\lim_{\Delta_1,\dots,\Delta_{n-1}\to\infty} \frac{\lambda^{\sigma}}{\lambda^{\sigma^*}} = 0$ for all $\sigma^* \in \mathcal{P}_{\sigma}^*$ and $\sigma \notin \mathcal{P}_{\sigma}^*$. Consequently, $\lim_{\Delta_1,\dots,\Delta_{n-1}\to\infty} \frac{D_1}{D_{1\infty}} = 1$, where $D_{1\infty} = \prod_{j=1}^m D_{1j}$ and

$$D_{11} = \det \begin{bmatrix} \binom{k_1}{1}\lambda_1^{-k_1} & \dots & \binom{k_1}{n_1-1}\lambda_1^{-k_1} \\ & \ddots & \\ \binom{k_{n_1-1}}{1}\lambda_1^{-k_{n_1-1}} & \dots & \binom{k_{n_1-1}}{n_1-1}\lambda_1^{-k_{n_1-1}} \end{bmatrix}$$
$$= \lambda_1^{-(k_1+\dots+k_{n_1-1})} \det \begin{bmatrix} \binom{k_1}{1} & \dots & \binom{k_1}{n_1-1} \\ & \ddots & \\ \binom{k_{n_1-1}}{1} & \dots & \binom{k_{n_1-1}}{n_1-1} \end{bmatrix}$$
$$= \left(\prod_{i=1}^{n_1-1} \frac{k_i}{i!} \prod_{1\le j<i\le n_1-1} (k_i-k_j)\right)\lambda_1^{-k(1)},$$

where the last equality follows from Lemma 20. Using Lemma 19, we can similarly derive that

$$D_{12} = \det \begin{bmatrix} \binom{k_{n_1}}{0}\lambda_2^{-k_{n_1}} & \dots & \binom{k_{n_1}}{n_2-1}\lambda_2^{-k_{n_1}} \\ & \ddots & \\ \binom{k_{n_1+n_2-1}}{0}\lambda_2^{-k_{n_1+n_2-1}} & \dots & \binom{k_{n_1+n_2-1}}{n_2-1}\lambda_2^{-k_{n_1+n_2-1}} \end{bmatrix}$$
$$= \lambda_2^{-(k_{n_1}+\dots+k_{n_1+n_2-1})} \det \begin{bmatrix} \binom{k_{n_1}}{0} & \dots & \binom{k_{n_1}}{n_2-1} \\ & \ddots & \\ \binom{k_{n_1+n_2-1}}{0} & \dots & \binom{k_{n_1+n_2-1}}{n_2-1} \end{bmatrix}$$
$$= \left(\prod_{i=0}^{n_2-1} \frac{1}{i!} \prod_{0\le j<i\le n_2-1} (k_{n_1+i}-k_{n_1+j})\right)\lambda_2^{-k(2)}.$$

Applying the same arguments to the rest of D_{1j}, we obtain that

$$D_{1\infty} = \prod_{i=1}^{n_1-1} \frac{k_i}{i!} \left(\prod_{v=2}^{m}\prod_{i=0}^{n_v-1} \frac{1}{v!} \prod_{0\le j<i\le n_v-1}\right.$$
$$\left. \times\ (k_{n_1+\dots+n_{v-1}+i} - k_{n_1+\dots+n_{v-1}+j})\right) \prod_{v=1}^m \lambda_v^{-k(v)}$$
$$\triangleq\ \psi_1(\{k_i\}_1^{n-1}) D_{\lambda}^1(\{k_i\}_1^{n-1}), \tag{44}$$

where $\psi_1(\{k_i\}_1^{n-1})$ is a multivariate polynomial in k_i and $D_{\lambda}^1(\{k_i\}_1^{n-1})$ contains k_i as its exponential component. Continuing with the same fashion, one can show that $\lim_{\Delta_1,\dots,\Delta_{n-1}\to\infty} \frac{D_2}{D_{2\infty}} = 1$ with $D_{2\infty} = \psi_2(\{k_i\}_1^{n-1}) D_{\lambda}^2(\{k_i\}_1^{n-1})$, where $\psi_2(\{k_i\}_1^{n-1})$ is a multivariate polynomial in k_i and

$$D_{\lambda}^2(\{k_i\}_1^{n-1}) = \frac{\prod_{v=3}^m \lambda_v^{-k(v)}}{\lambda_1^{k_1+\dots+k_{n_1}} \lambda_2^{k_{n_1+1}+\dots+k_{n_1+n_2-1}}}.$$

Then, it follows that

$$\lim_{\Delta_1,\dots,\Delta_{n-1}\to\infty} \frac{D_{2\infty}}{D_{1\infty}} = \left(\frac{\lambda_2}{\lambda_1}\right)^{k_{n_1}} \frac{\psi_2(\{k_i\}_1^{n-1})}{\psi_1(\{k_i\}_1^{n-1})} = 0$$

due to that $\psi_i(\{k_i\}_1^{n-1})$ is a multivariate polynomial in k_i and $|\lambda_1| > |\lambda_2|$.

A similar conclusion is reached for $D_3,\dots,D_m$, e.g.,

$$\lim_{\Delta_1,\dots,\Delta_{n-1}\to\infty} \frac{D_{v\infty}}{D_{1\infty}} = 0, \quad \forall v \ge 3.$$

To sum up, we finally get that

$$\lim_{\Delta_1,\dots,\Delta_{n-1}\to\infty} \frac{D_1+\dots+D_m}{D_{1\infty}} = 1.$$

By setting $\psi(\{k_i\}_1^{n-1}) = (\prod_{i=1}^m c_{i1}^{n_i})\psi_1(\{k_i\}_1^{n-1})$ and $D_{\lambda}(\{k_i\}_1^{n-1}) = (\prod_{i=1}^m \lambda_i^{n_i(n_i-1)/2}) D_{\lambda}^1(\{k_i\}_1^{n-1})$, it follows from (42) to (44) that

$$\lim_{\Delta_1,\dots,\Delta_{n-1}\to\infty} \frac{\det \mathcal{O}(\{k_i\}_1^{n-1})}{\psi(\{k_i\}_1^{n-1}) D_{\lambda}(\{k_i\}_1^{n-1})} = 1. \quad \square \tag{45}$$

Proof of Theorem 13. "$\Leftarrow$:" In order to simplify notation, we remove the dependence of $\{k_i\}_1^{n-1}$ for quantities in Lemma 21, i.e., rewrite $\mathcal{O}(\{k_i\}_1^{n-1})$ as $\mathcal{O}$. Then, it yields that

$$(\mathcal{O}^H\mathcal{O})^{-1} \leq \mathrm{tr}(\mathcal{O}^H\mathcal{O})^{-1}I = \sum_{i,j}\left(\frac{[\mathrm{adj}(\mathcal{O})]_{ij}}{\det(\mathcal{O})}\right)^2 I. \tag{46}$$

Here $\mathrm{adj}(\mathcal{O})$ is the adjoint matrix of $\mathcal{O}$ and $[\mathrm{adj}(\mathcal{O})]_{ij}$ is the (i,j)-th element of $\mathrm{adj}(\mathcal{O})$. Following a similar line of Lemma 21, we can show that there exist constant numbers $\beta_{i,j} = \beta_{i,j}(\lambda_1,\ldots,\lambda_m)$ such that for sufficiently large Δ_j, we have that $\mathrm{adj}(\mathcal{O})_{ij} \leq \beta_{i,j}|\psi|\prod_{v=1}^m|\lambda_v|^{-k'(v)}$, where $k'(v) = k_1+\cdots+k_{n_1-2}$ and $k'(v) = k_{n_1+\cdots+n_{v-1}-1}+\cdots+k_{n_1+\cdots+n_v-2}$ if $v \geq 2$. In light of (46), it follows that there exist constant numbers $\tilde{\beta}_{ij} = \tilde{\beta}_{ij}(\lambda_1,\ldots,\lambda_m)$ such that

$$\limsup_{\Delta_1,\ldots,\Delta_{n-1}\to\infty} \frac{(\mathcal{O}^H\mathcal{O})^{-1}}{\prod_{i=1}^m|\lambda_v|^{2\Delta(v)}} \leq \left(\sum_{i,j}\tilde{\beta}_{i,j}\right)I, \tag{47}$$

where $\Delta(1) = \Delta_1+\cdots+\Delta_{n_1-1}$ and $\Delta(v) = \Delta_{n_1+\cdots+n_{v-1}}+\cdots+\Delta_{n_1+\cdots+n_v-1}$, $v \geq 2$. The rest of the proof directly follows from that of Theorem 12.

"$\Rightarrow$:" It is proved in Theorem 3. $\square$

References

Ash, R., & Doléans-Dade, C. (2000). *Probability and measure theory*. Academic Press.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization*, *31*, 942–959.

Censi, A. (2011). Kalman filtering with intermittent observations: convergence for semi-Markov chains and an intrinsic performance measure. *IEEE Transactions on Automatic Control*, *56*(2), 376–381.

Dey, S., Leong, A., & Evans, J. (2009). Kalman filtering with faded measurements. *Automatica*, *45*(10), 2223–2233.

Epstein, M., Shi, L., Tiwari, A., & Murray, R. (2008). Probabilistic performance of state estimation across a lossy network. *Automatica*, *44*(12), 3046–3053.

Gupta, V., Dana, A., Hespanha, J., Murray, R., & Hassibi, B. (2009). Data transmission over networks for estimation and control. *IEEE Transactions on Automatic Control*, *54*(8), 1807–1819.

Gupta, V., Martins, N., & Baras, J. (2009). Optimal output feedback control using two remote sensors over erasure channels. *IEEE Transactions on Automatic Control*, *54*(7), 1463–1476.

Hespanha, J., Naghshtabrizi, P., & Xu, Y. (2007). A survey of recent results in networked control systems. *Proceedings of the IEEE*, *95*(1), 138–162.

Horn, R., & Johnson, C. (1985). *Matrix analysis*. Cambridge University Press.

Hu, S., & Yan, W. (2007). Stability robustness of networked control systems with respect to packet loss. *Automatica*, *43*(7), 1243–1248.

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

Kar, S., Sinopoli, B., & Moura, J. (2010). Kalman filtering with intermittent observations: weak convergence to a stationary distribution. *IEEE Transactions on Automatic Control*, accepted by.

Kluge, S., Reif, K., & Brokate, M. (2010). Stochastic stability of the extended Kalman filter with intermittent observations. *IEEE Transactions on Automatic Control*, *55*(2), 514–518.

Meyn, S., Tweedie, R., & Hibey, J. (1996). *Markov chains and stochastic stability*. London: Springer-Verlag.

Mo, Y., & Sinopoli, B. 2010, Towards finding the critical value for Kalman filtering with intermittent observations, http://arxiv.org/abs/1005.2442.

Mostofi, Y., & Murray, R. (2009). To drop or not to drop: design principles for Kalman filtering over wireless fading channels. *IEEE Transactions on Automatic Control*, *54*(2), 376–381.

Plarre, K., & Bullo, F. (2009). On Kalman filtering for detectable systems with intermittent observations. *IEEE Transactions on Automatic Control*, *54*(2), 386–390.

Schenato, L. (2008). Optimal estimation in networked control systems subject to random delay and packet drop. *IEEE Transactions on Automatic Control*, *53*(5), 1311–1317.

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

Shi, L., Epstein, M., & Murray, R. (2010). Kalman filtering over a packet-dropping network: a probabilistic perspective. *IEEE Transactions on Automatic Control*, *55*(3), 594–604.

Singer, R. (1970). Estimating optimal filter tracking performance for manned maneuvering targets. *IEEE Transactions on Aerospace and Electronic Systems*, *6*(4), 473–483.

Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M., & Sastry, S. (2004). Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, *49*(9), 1453–1464.

Solo, V. 1991, One step ahead adaptive controller with slowly time-varying parameters, Technical report, Dept. ECE, John Hopkins University, Baltimore.

Sun, S., Xie, L., Xiao, W., & Soh, Y. (2008). Optimal linear estimation for systems with multiple packet dropouts. *Automatica*, *44*(5), 1333–1342.

Trivellato, M., & Benvenuto, N. (2010). State control in networked control systems under packet drops and limited transmission bandwidth. *IEEE Transactions on Communications*, *58*(2), 611–622.

Xiao, N., Xie, L., & Fu, M. (2009). Kalman filtering over unreliable communication networks with bounded Markovian packet dropouts. *International Journal of Robust and Nonlinear Control*, *19*(16), 1770–1786.

Xie, L., & Xie, L. (2008). Stability of a random Riccati equation with Markovian binary switching. *IEEE Transactions on Automatic Control*, *53*(7), 1759–1764.

Xie, L., & Xie, L. (2009). Stability analysis of networked sampled-data linear systems with markovian packet losses. *IEEE Transactions on Automatic Control*, *54*(6), 1368–1374.

You, K., Fu, M., & Xie, L. 2011, Necessary and suffficient conditions for stability of Kalman filtering with Markovian packet losses, In: Proceedings of 18th IFAC World Congress, Milano, Italy, pp. 12465–12470.

You, K., & Xie, L. (2010). Minimum data rate for mean square stabilization of discrete LTI systems over lossy channels. *IEEE Transactions on Automatic Control*, *55*(10), 2373–2378.

You, K., & Xie, L. (2011). Minimum data rate for mean square stabilizability of linear systems with Markovian packet losses. *IEEE Transactions on Automatic Control*, *56*(4), 772–785.

Keyou You was born in Jiangxi Province, China, in 1985. He received the B.S. degree in statistical science from Sun Yat-sen (Zhongshan) University, Guangzhou, China, in 2007 and is currently pursuing the Ph.D. degree in electrical and electronic engineering at Nanyang Technological University, Singapore.

From May 2010 to July 2010, he was with the ARC Center of Excellence for Complex Dynamic Systems and Control, the University of Newcastle, Australia, as a visiting scholar. His current research interests include quantized estimation and control, networked control, system identification and distributed control and estimation.

He won the Guan Zhaozhi best paper award at the 29th Chinese Control Conference, Beijing, China, 2010.

Minyue Fu received his Bachelor's Degree in Electrical Engineering from the University of Science and Technology of China, Hefei, China, in 1982, and M.S. and Ph.D. degrees in Electrical Engineering from the University of Wisconsin-Madison in 1983 and 1987, respectively. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan. He joined the Department of Electrical and Computer Engineering, the University of Newcastle, Australia, in 1989. Currently, he is a Chair Professor in Electrical Engineering. He has served as a Head of School at the University of Newcastle. In addition, he was a Visiting Associate Professor at University of Iowa in 1995–1996, a Senior Fellow/Visiting Professor at Nanyang Technological University, Singapore, 2002, and a Visiting Professor at Tokyo University in 2003. He has held a ChangJiang Visiting Professorship at Shandong University, China. Currently, he holds a Qian-ren Professorship at Zhejiang University, China.

He is a Fellow of IEEE. His main research interests include control systems, signal processing and communications. He has been an Associate Editor for the IEEE Transactions on Automatic Control, Automatica and Journal of Optimization and Engineering.

Lihua Xie received the B.E. and M.E. degrees in electrical engineering from Nanjing University of Science and Technology in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Newcastle, Australia, in 1992. He held teaching appointments in the Department of Automatic Control, Nanjing University of Science and Technology from 1986 to 1989. Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a professor and Head of Division of Control and Instrumentation. He is a Changjiang visiting professor with South China University of Technology.

Dr Xie's research interests include robust control and estimation, networked control systems, sensor networks, time delay systems and control of hard disk drive systems. He is an editor of IET Book Series on Control and has served as an Associate Editor of several journals including IEEE Transactions on Automatic Control, Automatica, IEEE Transactions on Control Systems Technology, IEEE Transactions on Circuits and Systems-II, and IET Proceedings on Control Theory and Applications. Dr Xie is a Fellow of IEEE and a Fellow of IFAC.