Kalman Filtering Over Fading Channels: Zero–One Laws and Almost Sure Stabilities

Junfeng Wu, Guodong Shi, Brian D. O. Anderson, *Life Fellow, IEEE*, and Karl Henrik Johansson

***Abstract*— In this paper, we investigate probabilistic stability of Kalman filtering over fading channels modeled by $*$-mixing random processes, where channel fading is allowed to generate non-stationary packet dropouts with temporal and/or spatial correlations. Upper/lower almost sure (a.s.) stabilities and absolutely upper/lower a.s. stabilities are defined for characterizing the sample-path behaviors of the Kalman filtering. We prove that both upper and lower a.s. stabilities follow a zero–one law, i.e., these stabilities must happen with a probability either zero or one, and when the filtering system is one-step observable, the absolutely upper and lower a.s. stabilities can also be interpreted using a zero–one law. We establish general stability conditions for (absolute) upper and lower a.s. stabilities. In particular, with one-step observability, we show the equivalence between absolutely a.s. stabilities and a.s. ones, and necessary and sufficient conditions in terms of packet arrival rate are derived; for the so-called non-degenerate systems, we also manage to give a necessary and sufficient condition for upper a.s. stability.**

***Index Terms*— Kalman filter, fading channels, stability.**

I. Introduction

A. Background and Motivation

THE last decade has witnessed an increasing attention on wireless sensor networks (WSNs) from the control, communication and networking communities, thanks to a rapid development of micro–electronics, wireless communication, and information and networking technologies. WSNs have applications in a wide range of areas such as health care, intelligent buildings, smart transportation and power grid, just to name a few, due to considerable advantages, including reducing operational cost, allowing distributed sensing and information sharing among different nodes, etc. New challenges have also been introduced at the expense of the aforementioned advantages, where control and estimation systems have to be sustainable in the presence of communication links. This has attracted significant attention to the study of information theory for network systems [2], and one fundamental aspect lies in that channel fading [3] leads to constructive or destructive interference of telecommunication signals, and at times severe drops in the channel signal–to–noise ratio may cause temporary communication outage for the underlying control or estimation systems.

The Kalman filter [4], [5] plays a fundamental role in networked state estimation systems, where a basic theme is the stability of Kalman filtering over a communication channel between the plant and the estimator which generates random packet dropouts [6]. There were mainly two stability categories in the literature focusing on the mean–square, or the probability distribution, evolution of the error covariance along sample–paths of the Kalman filtering, respectively. The majority of the research works assumes the channel admits identically and independently distributed (i.i.d.) or Markovian packet drops. Sinopoli *et al.* [7] modeled the packet losses as an i.i.d. Bernoulli process, and proved that there exists a critical arrival rate for the packet arrival rate, below which, the expected prediction error covariance is unbounded. Further improvements of this result were developed in [8]–[10]. The mean–square stability, and stability defined at random packet recovery/reception times, of Kalman filtering subject to Markovian packet losses generated by a Gilbert–Elliott channel were studied in [11]–[16]. Efforts have also been made from a probabilistic point of view. Weak convergence of Kalman filtering with intermittent observations, which amounts to having the error covariance matrix converge to a limit distribution, were investigated in [17]–[19] for i.i.d., semi–Markov, and Markovian packet drop models, respectively. The weak convergence of distributed Kalman filtering was studied in [20].

In this paper, we aim to characterize the asymptotic behaviors of the sample paths of Kalman filtering over fading channels. Instead of only focusing on certain average property (mean–square, or distribution) of the sample paths, we go beyond most of the stability notions considered in the literature. It turns out that the majority of the packet drop models can be put under a unified model from the mixing theory.

B. Model and Contribution

We assume that the data packets are regarded as successfully received when received error–free; and are regarded as

Manuscript received December 27, 2014; revised November 24, 2017 and July 28, 2018; accepted July 30, 2018. Date of publication August 13, 2018; date of current version September 13, 2018. This work was supported in part by the Knut and Alice Wallenberg Foundation, in part by the Swedish Foundation for Strategic Research, in part by the Swedish Research Council, VINNOVA, in part by the NNSF of China under Grant 61790570, Grant 61790573, and Grant 61120106011, and in part by the Australian Research Council under Grant DP-130103610. This paper was presented at the 53rd 2014 IEEE Conference on Decision and Control [1].

J. Wu is with the College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: jfwu@zju.edu.cn).

G. Shi is with the Research School of Engineering, The Australian National University, Canberra, ACT 0200, Australia (e-mail: guodong.shi@anu.edu.au).

B. D. O. Anderson is with the Hangzhou Dianzi University, Hangzhou 310000, China, and also with the Research School of Engineering, Australian National University, Canberra, ACT 0200, Australia (e-mail: brian.anderson@anu.edu.au).

K. H. Johansson is with the ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: kallej@kth.se).

Communicated by P. Mitran, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2018.2865381

completely lost otherwise. Although real digital communication introduces a bunch of other challenges, such as quantization and data rate, bit errors, and random delays [2], we are exclusively devoted to studying the impact of packet dropouts on the estimation performance and therefore those other effects will be ignored. To address non–stationarity of the propagation environment with spatial and temporal correlations between channel parameters [21]–[23], we introduce a packet drop process that is $*$-mixing [24]. The mixing theory provides a tool of investigating random processes which are approximately independent in the sense that the dependence dies away as the distance of any two random variables in the process grows large. The $*$-mixing model includes but also generalizes i.i.d. and Markov–type models in the literature.

We consider the probabilistic stabilities of Kalman filtering over such general fading channels. We devise the definitions of upper/lower a.s. stabilities and absolutely upper/lower a.s. stabilities. The difference and connection between mean–square stability and (absolutely) a.s. stabilities are also discussed. Consistent with a.s. convergence, the definitions of (absolutely) a.s. stabilities serve as a supplement of the stability study on Kalman filtering from the perspective of probabilistic behaviors. We establish the following results:

- We prove that the upper and lower a.s. stabilities follow a zero–one law, indicating that an event must happen with probability either zero or one. When the considered filtering system is one–step observable, the absolutely upper and lower a.s. stabilities can also be interpreted by the zero–one law.
- We further present stability conditions for the (absolutely) upper and lower stabilities. We first give sufficient/necessary conditions for general linear time–invariant (LTI) systems. One–step observable systems yield tighter results with necessary and sufficient conditions in terms of the packet arrival rate derived for upper and lower a.s. stabilities. It is also shown for one–step observable systems that a.s. stability is equivalent to absolutely a.s. one. Finally, for the so–called non–degenerate systems, we manage to give a necessary and sufficient upper a.s. stability condition.

All the above results are established under $*$-mixing fading channels, and to the best of our knowledge, this is the first time the concept of mixing has been introduced to the modelling of random packet losses. An embryo of part of this work (some stability conditions) was presented in [1] for independent channels.

C. Paper Organization

The remainder of the paper is organized as follows. Section II provides the problem setup, defines the (absolutely) upper/lower a.s. stabilities, and introduces the $*$-mixing random process considered in [24]. The difference between various stabilities are also discussed in Section II. In Section III, two stability zero–one laws are derived. Various stability conditions are studied in Section IV. Some concluding remarks are given in the end.

Notations: $\mathbb{N}$ is the set of positive integers. For a real number x, $\lfloor x \rfloor$ and $\lceil x \rceil$ denote the largest integer not greater than x and the smallest integer not less than x respectively. The set of n by n symmetric positive semi–definite (positive definite) matrices over the complex field is denoted as $\mathbb{S}_+^n$ ($\mathbb{S}_{++}^n$). For a matrix X, X denotes the transpose of X and X^* the conjugate transpose of X. Moreover, $\lambda_i(X)$ represents the ith largest eigenvalue of X in terms of magnitude for $i = 1, \dots, n$, and $\|X\|_2$ represents the spectral norm of X. The indicator function of a subset $\mathcal{A} \subset \Omega$ is a function $1_{\mathcal{A}} : \Omega \to \{0, 1\}$, where $1_{\mathcal{A}}(\omega) = 1$ if $\omega \in \mathcal{A}$, otherwise $1_{\mathcal{A}}(\omega) = 0$. $\sigma(\cdot)$ denotes the σ–algebra generated by random variables. For an event $\mathcal{A}$ in some probability space, "$\mathcal{A}$ $i.o.$" means $\mathcal{A}$ happens infinitely often.

II. Kalman Filtering Over Fading Channels

In this section, we introduce the Kalman filtering model and define the problem of interest.

A. Kalman Filtering With Packet Dropouts

Consider an LTI system:

$$x_{k+1} = Ax_k + w_k, \tag{1}$$

$$y_k = Cx_k + v_k, \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the process state vector, $y_k \in \mathbb{R}^m$ is the observation vector, $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are zero–mean Gaussian random vectors with $\mathbb{E}[w_k w_j{}'] = \delta_{kj} Q$ ($Q \geq 0$), $\mathbb{E}[v_k v_j{}'] = \delta_{kj} R$ ($R > 0$), and $\mathbb{E}[w_k v_j{}'] = 0\ \forall j, k$. The δ_{kj} is the Kronecker delta function with $\delta_{kj} = 1$ if $k = j$ and $\delta_{kj} = 0$ otherwise. The initial state x_0 is a zero–mean Gaussian random vector that is uncorrelated with w_k and v_k and has covariance $P_0 \geq 0$. We assume that the pair (C, A) is observable and $(A, Q^{1/2})$ controllable. We introduce the standard definition of observability index of the pair (C, A).

Definition 1: For the observable pair (C, A), the *observability index* $\mathrm{I_o} \in \mathbb{N}$ is defined as the smallest integer such that $[C', A'C', \dots, (A^{\mathrm{I_o}-1})'C']'$ has full column rank.
It is evident that $\mathrm{I_o} \leq n$.

Purely stable LTI systems do not interest us as their estimation error covariance matrix automatically decays. In that case, it becomes trivial to discuss probabilistic stability issues. For unstable LTI systems, it can be seen that, by applying a similarity transformation, the unstable and stable modes can be decoupled. An open–loop prediction for the stable mode always has a bounded estimation error covariance, therefore, this mode does not play any key role in the problem considered here. Without loss of generality, we assume that

(A1) *All of the eigenvalues of A have magnitudes no less than 1.*

We consider an estimation scheme where the raw measurements $\{y_k\}_{k\in\mathbb{N}}$ of the sensor are transmitted to the estimator over an erasure communication channel over which packets may be dropped randomly, see Fig. 1. We assume that the packets are regarded as successfully received when received error–free; and are regarded as completely lost otherwise. Denote by $\gamma_k \in \{0, 1\}$ the arrival of y_k at time k:

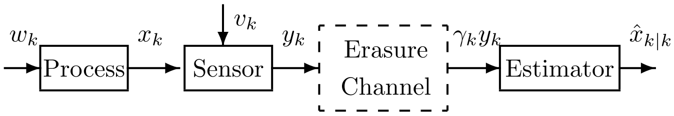

Fig. 1. State estimation over an erasure channel.

If $\gamma_k = 1$, it indicates that y_k successfully arrives at the estimator; otherwise $\gamma_k = 0$. We assume that the sequence $\{\gamma_k\}_{k\in\mathbb{N}}$ is independent of how the system evolves, and that the estimator knows whether the packet has arrived or not at each time. Define $\mathcal{F}_k$ as the filtration generated by all the measurements received by the estimator up to time k, i.e., $\mathcal{F}_k \triangleq \sigma(\gamma_t y_t, \gamma_t; 1 \le t \le k)$, and define $\mathcal{F} = \sigma\left(\cup_{k=1}^{\infty}\mathcal{F}_k\right)$. We use a triple $(\Omega, \mathcal{F}, \mathbb{P})$ to denote the common probability space for all random elements in the LTI system as well as in the packet dropouts. The estimator computes $\hat{x}_{k|k}$, the minimum mean–squared error estimate, and $\hat{x}_{k+1|k}$, the one–step prediction, according to $\hat{x}_{k|k} = \mathbb{E}[x_k|\mathcal{F}_k]$ and $\hat{x}_{k+1|k} = \mathbb{E}[x_{k+1}|\mathcal{F}_k]$, where $\mathbb{E}$ denotes the expectation induced by $\mathbb{P}$. Let $P_{k|k}$ and $P_{k+1|k}$ be the corresponding estimation and prediction error covariance matrices, receptively, i.e., $P_{k|k} = \mathbb{E}[(x_k - \hat{x}_{k|k})(\cdot)'|\mathcal{F}_k]$ and $P_{k+1|k} = \mathbb{E}[(x_{k+1} - \hat{x}_{k+1|k})(\cdot)'|\mathcal{F}_k]$, which are computed recursively via a modified Kalman filter [7]:

$$\begin{aligned}
K_k &= P_{k|k-1}C'(CP_{k|k-1}C' + R)^{-1},\\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + \gamma_k K_k(y_k - CA\hat{x}_{k|k-1}),\\
P_{k|k} &= (I - \gamma_k K_k C)P_{k|k-1},\\
\hat{x}_{k+1|k} &= A\hat{x}_{k|k},\\
P_{k+1|k} &= AP_{k|k}A' + Q.
\end{aligned}$$

In particularly, $P_{k+1|k}$ evolves in the following way

$$\begin{aligned}
P_{k+1|k} = {}& AP_{k|k-1}A' + Q\\
&- \gamma_k AP_{k|k-1}C'(CP_{k|k-1}C' + R)^{-1}CP_{k|k-1}A'. \quad (3)
\end{aligned}$$

It can be seen that $P_{k+1|k}$ now becomes a function of the random variables $\{\gamma_t\}_{1\le t\le k}$. In what follows, we are devoted to characterizing the impacts of $\{\gamma_k\}_{k\in\mathbb{N}}$ on $P_{k+1|k}$. To simplify discussion in the sequel, let us use a simpler notation $P_{k+1} \triangleq P_{k+1|k}$, and introduce the functions h, g, h^k and g^k: $\mathbb{S}_+^n \to \mathbb{S}_+^n$ as follows:

$$h(X) \triangleq AXA' + Q, \quad (4)$$

$$g(X) \triangleq AXA' + Q - AXC'(CXC' + R)^{-1}CXA', \quad (5)$$

$$h^k(X) \triangleq \underbrace{h \circ h \circ \cdots \circ h}_{k \text{ times}}(X) \text{ and } g^k(X) \triangleq \underbrace{g \circ g \circ \cdots \circ g}_{k \text{ times}}(X),$$

where $\circ$ denotes the function composition.

B. $*$-Mixing Fading Channels

Wireless channels are mainly affected by path loss, small–scale fading and shadow fading. In a wireless connected vehicle–to–vehicle network [23], for example, for the sake of moving vehicles, small–scale fading happens in an unpredictable way. Moreover, shadow fading, caused by obstructing objects, leads to temporal and spatial correlations between communications links. The aforementioned factors are no longer negligible. To model packet dropouts subject to spatially and/or temporally correlated and non–stationary fading channels, on one hand, we need to take the non–stationarity of propagation environment and correlations between channel parameters into account; on the other hand, we have to retain indispensable assumptions, making it possible to build up instructive theories upon it. We model the packet dropouts as a $*$-mixing stochastic process, where the concept of mixing, originating from physics, is an attempt to interpret the thermodynamic behavior of mixtures.

Before proceeding, we introduce the definition of $*$-mixing, which is taken from [24].

Definition 2: The sequence of random variables $\{\xi_k\}_{k\in\mathbb{N}}$ on a probability space $(\mathscr{S}, \mathcal{S}, \mu)$ is said to be $*$-mixing if there exists a positive integer N and a real–valued function f defined for $n \ge \mathrm{N}$, where $n \in \mathbb{N}$, such that

(i) f is a non–increasing function with $\lim_{n\to\infty} f(n) = 0$;
(ii) There holds $\left|\mu(\mathcal{A}\cap\mathcal{B}) - \mu(\mathcal{A})\mu(\mathcal{B})\right| \le f(n)\mu(A)\mu(B)$ for all $n \ge \mathrm{N}$, $\mathcal{A} \in \sigma(\xi_1, \ldots, \xi_k), \mathcal{B} \in \sigma(\xi_{k+n}, \xi_{k+n+1}, \ldots)$, and $k \in \mathbb{N}$.

In the sequel, we assume that

(A2) *The random process $\{\gamma_k\}_{k\in\mathbb{N}}$ is $*$-mixing.*

To the best of our knowledge, this is the first time mixing has been introduced when modelling random packet dropouts. One coarse way to explain the above mathematical definition is that $*$-mixing implies that the occurrence of any two groups of possible states can be considered approximately independent as long as the two groups are a sufficient amount of time apart from each other, where dependence is "quantified" by the mixing coefficient $f(n)$. It is a universal understanding that in the physical world historical states in remote past impact less and less on the evolution of future states, provided that the hypothesis of $*$-mixing stands. Note that the idea of mixing has been used in the "theoretical channel model" (the theoretical channel refers to a mapping from the input source to the output source) in the literature [25]–[27].

Remarkably enough the mixing model admits most of the well–studied models reported in the literature, e.g., i.i.d. [7], [9], [17], Markov [10], [14], [15], [19], semi–Markov [16], [18], Markovian jump [13], finite–state Markov [28], as its special cases [24].

C. Problems of Interest

In this paper, we are interested in the sample–path behaviors of Kalman filtering with $*$-mixing packet losses. Since $\mathrm{Tr}(P_k)$ represents the sum of squared error variance of the estimate for each element of x_k, we use $\mathrm{Tr}(P_k)$ as a performance metric. Noting that $\limsup_{k\to\infty}\mathrm{Tr}(P_k)$ and $\liminf_{k\to\infty}\mathrm{Tr}(P_k)$ are well–defined random variables taking values from $\mathbb{R}\cup\{+\infty\}$, we introduce the following stability notions for the considered Kalman filter.

Definition 3: The considered Kalman filter is termed

(i) upper a.s. stable if $\mathbb{P}\left(\limsup_{k\to\infty}\mathrm{Tr}(P_k) < \infty\right) = 1$, and lower a.s. stable if $\mathbb{P}\left(\liminf_{k\to\infty}\mathrm{Tr}(P_k) < \infty\right) = 1$;
(ii) absolutely upper a.s. stable if there exists a constant $\mathrm{C} > 0$ such that $\mathbb{P}\left(\limsup_{k\to\infty}\mathrm{Tr}(P_k) < \mathrm{C}\right) = 1$,

and absolutely lower a.s. stable if there exists a constant $\mathrm{C} > 0$ such that $\mathbb{P}\big(\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \mathrm{C}\big) = 1$.

These stability notions focus on the asymptotic behavior of the estimation system along every sample path across the sample space, enabling us to investigate a Kalman filtering system from a probabilistic perspective. Note that, in general, absolutely a.s. stability is a stronger notion than the a.s. one. For convenience, we also call the considered Kalman filter *upper a.s. unstable* if $\mathbb{P}\big(\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \infty\big) = 0$, and *lower a.s. unstable* if $\mathbb{P}\big(\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \infty\big) = 0$. Additionally, the Kalman filter is said to be almost surely convergent if

$$\mathbb{P}\Big(\lim_{k\to\infty} P_k \text{ exists, and is finite}\Big) = 1. \tag{6}$$

D. Discussions

In the literature, a widely investigated stability notion of Kalman filtering systems with packet losses is mean–square stability, i.e., the Kalman filtering is mean–square stable if $\sup_{k\in\mathbb{N}} \mathbb{E}\|P_k\| < \infty$. In general, there are no implications between a.s./absolutely a.s. stabilities and mean–square stability for the Kalman filter. This relation is analogous to the relation between a.s. convergence and L_p–convergence for a sequence of random variables [29], because a.s./absolutely a.s. stabilities are defined on the basis of a.s. convergence and mean–square stability is defined on the basis of L_p–convergence.

Another important concept of Kalman filtering systems is the weak–convergence, which requires P_k to converge to a limit in distribution [17]–[19]. Then by standard chain of implications of the notions of probabilistic convergence, we know that both the mean–square convergence (i.e., $\lim_{k\to\infty} \mathbb{E}\|P_k - P_*\| = 0$ for some P_*) and the a.s. convergence (6) imply weak convergence.

III. The Zero–One Laws

A tail event of a random process is an event whose occurrence is independent of each finite subset of random variables. In this section, we present that the a.s. stabilities follow a zero–one law, which is shown with the aid of the definition of tail events and the zero–one law for a $*$-mixing sequence.

Theorem 1: Let Assumptions (A1)–(A2) hold. Both upper and lower a.s. stabilities follow a zero–one law, i.e.,

(i) Either $\mathbb{P}\big(\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \infty\big) = 1$ or $\mathbb{P}\big(\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \infty\big) = 0$;

(ii) Either $\mathbb{P}\big(\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \infty\big) = 1$ or $\mathbb{P}\big(\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \infty\big) = 0$.

The following theorem further shows that the zero–one law also applies to absolutely upper and lower a.s. stabilities when the system is one–step observable, i.e., $\mathrm{I_o} = 1$.

Theorem 2: Let Assumptions (A1)–(A2) hold. Suppose $\mathrm{I_o} = 1$. Then absolutely upper and lower a.s. stabilities follow the zero–one law, i.e.,

(i) Either there exists a constant $\mathrm{C} > 0$ such that $\mathbb{P}\big(\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \mathrm{C}\big) = 1$ or $\mathbb{P}\big(\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \mathrm{C}\big) = 0$ holds for any $\mathrm{C} > 0$;

(ii) Either there exists a constant $\mathrm{C} > 0$ such that $\mathbb{P}\big(\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \mathrm{C}\big) = 1$ or $\mathbb{P}\big(\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \mathrm{C}\big) = 0$ holds for any $\mathrm{C} > 0$.

In the rest of this section, we first gather and establish a few supporting lemmas, and then provide detailed proofs for Theorems 1 and 2.

A. Supporting Lemmas

Denote the unique solution to $g(X) = X$ as $\overline{P}$. Assuming the observability of (C, A) and controllability of $(A, Q^{1/2})$, it is well known that $\overline{P}$ is a positive definite matrix [30]; and that, for a standard Kalman filter, $\lim_{k\to\infty} P_k = \overline{P}$ [31]. For the operators h and g, the following lemma holds. The proof can be found in [10, Lemma A.1].

Lemma 1: For any matrices $X \geq Y \geq 0$,

$$h(X) \geq h(Y), \tag{7}$$

$$g(X) \geq g(Y), \tag{8}$$

$$h(X) \geq g(X). \tag{9}$$

The following two lemmas further establish some useful properties of operators g and h.

Lemma 2: For any $X \in \mathbb{S}_+^n$, there exists an integer $t \in \mathbb{N}$, independent of X, such that $g^t(X) > 0$.

Proof: Choose a constant $\beta \in (0, 1)$. Since $\lim_{k\to\infty} g^k(0) = \overline{P}$, there always exists a sufficiently large integer $\mathrm{N}(\beta)$ such that $g^k(0) \geq \beta\overline{P}$ for all $k \geq \mathrm{N}(\beta)$. Then By Lemma 1, $g^k(X) \geq g^k(0) > 0$. Note that $\mathrm{N}(\beta)$ is chosen independent of X. The conclusion follows by letting $t = \mathrm{N}(\beta)$. □

Lemma 3: There exists a constant $a > 0$ such that $\mathrm{Tr}\left(h^k(X)\right) \geq a|\lambda_1(A)|^{2k}$ holds for all $X \in \mathbb{S}_+^n$ and for all $k \in \mathbb{N}$.

Proof: By the controllability of $(A, Q^{1/2})$ assumed, one has $V \triangleq h^n(0) > 0$. Then there always exists a real number $a_0 > 0$ so that $V \geq a_0 I$. Therefore, $h^k(0) \geq a_0\, A^{k-n}(A')^{k-n}$ holds for all $k \geq n$. Let us denote the Schur's unitary triangularization [32] of A as $A = UTU^*$ where U is a unitary matrix and $T = [t_{ij}]$ is an upper triangular with $t_{ii} = \lambda_i(A)$, $i = 1, \ldots, n$. Since $A^{k-n}(A')^{k-n}$ is Hermitian and positive semi–definite, $\lambda_1\left(A^{k-n}(A')^{k-n}\right)$ is real and moreover,

$$\begin{aligned}
\lambda_1\left(A^{k-n}(A')^{k-n}\right) &= \lambda_1\left(T^{k-n}(T^*)^{k-n}\right) \\
&= \left\| \begin{bmatrix} \lambda_1(A^{k-n}) & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n(A^{k-n}) \end{bmatrix} \right\|_2^2 \\
&\geq |\lambda_1(A^{k-n})|^2 \\
&= |\lambda_1(A)|^{2(k-n)}.
\end{aligned}$$

Therefore, $\mathrm{Tr}\left(h^k(0)\right) \geq a_n|\lambda_1(A)|^{2k}$ holds for all $k \geq n$ with $a_n \triangleq a_0|\lambda_1(A)|^{-2n}$. As for $k = 1, \ldots, n-1$, we choose a sequence of positive real numbers, denoted by $\{a_k\}_{1\leq k\leq n-1}$, such that $\mathrm{Tr}\left(h^k(0)\right) \geq a_k|\lambda_1(A)|^{2k}$. The conclusion follows by taking $a \triangleq \min\{a_k : k = 1, \ldots, n\} > 0$. □

Since (C, A) is observable, $J \triangleq \left[(CA^{\mathrm{I_o}-1})', (CA^{\mathrm{I_o}-2})', \ldots, C'\right]'$ has full column rank and $J'J$ is nonsingular. Denote

$$M_0 \triangleq (J'J)^{-1}J'\left(H\begin{bmatrix} Q & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Q \end{bmatrix}H' + \begin{bmatrix} R & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & R \end{bmatrix}\right) \times J(J'J)^{-1} \quad (10)$$

and

$$H = \begin{bmatrix} C & CA & \dots & \dots & CA^{\mathrm{I_o}-2} \\ 0 & C & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & C & CA \\ 0 & 0 & 0 & \dots & C \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Next define

$$M = h^{\mathrm{I_o}}(M_0). \quad (11)$$

For $\mathrm{I_o}$ and M, we have the following lemma.

Lemma 4: Suppose that by time $k-1$ there are at least $\mathrm{I_o}$ consecutive measurements $y_{k-\mathrm{I_o}}, \dots, y_{k-1}$ received by the Kalman filter. Then there holds $P_k \le M$.

Proof: Observe that

$$\begin{bmatrix} y_{k-1} \\ y_{k-2} \\ \vdots \\ y_{k-\mathrm{I_o}} \end{bmatrix} = Jx_{k-\mathrm{I_o}} + H\begin{bmatrix} w_{k-2} \\ w_{k-3} \\ \vdots \\ w_{k-\mathrm{I_o}} \end{bmatrix} + \begin{bmatrix} v_{k-1} \\ v_{k-2} \\ \vdots \\ v_{k-\mathrm{I_o}} \end{bmatrix}.$$

Based on the consecutive measurements $y_{k-\mathrm{I_o}}, \dots, y_{k-1}$ received by the estimator, we use the following estimator to generate a linear prediction of x_k:

$$\bar{x}_k = A^{\mathrm{I_o}}(J'J)^{-1}J'\begin{bmatrix} y_{k-1} \\ y_{k-2} \\ \vdots \\ y_{k-\mathrm{I_o}} \end{bmatrix}.$$

The associated prediction error covariance $\overline{P}_k \triangleq \mathbb{E}[(x_k - \bar{x}_k)(x_k - \bar{x}_k)'] \le M$. Since the Kalman filter is known as the linear minimum mean–squared error estimator, we have $P_k \le \overline{P} \le M$, which completes the proof. □

In the following, we introduce the definition of Riemannian distance on $\mathbb{S}^n_{++}$.

Definition 4: For any $X, Y \in \mathbb{S}^n_{++}$, the Riemannian distance δ between X and Y is defined as

$$\delta(X,Y) = \left(\sum_{i=1}^{n} \mathrm{Log}^2\lambda_i\left(XY^{-1}\right)\right)^{1/2}. \quad (12)$$

It has been shown that δ is a metric on $\mathbb{S}^n_{++}$, and that the metric space $(\mathbb{S}^n_{++}, \delta)$ is complete and separable [33]. In $(\mathbb{S}^n_{++}, \delta)$, the operators $h,\ g$ defined in (4) and (5) are non–expansive and $g^{\mathrm{I_o}}$ is contractive.

Lemma 5 (34, Th. 1.7): Suppose that A is invertible. In the metric space $(\mathbb{S}^n_{++}, \delta)$,

(i) There hold $\delta(h(X), h(Y)) \le \delta(X,Y)$ and $\delta(g(X), g(Y)) \le \delta(X,Y)$ for any $X, Y \in \mathbb{S}^n_{++}$;

(ii) There exists a real number $q \in (0,1)$ that only depends on $A,\ C,\ Q,\ R$ such that there holds

$$\delta(g^{\mathrm{I_o}}(X), g^{\mathrm{I_o}}(Y)) \le q\delta(X,Y)$$

for any $X, Y \in \mathbb{S}^n_{++}$.

It is also easy to establish the following lemma.

Lemma 6: In the metric space $(\mathbb{S}^n_{++}, \delta)$, there holds $2^{-\delta(X,Y)}X \le Y \le 2^{\delta(X,Y)}X$ for any $X, Y \in \mathbb{S}^n_{++}$.

Proof: From the definition of Riemannian distance in (12), we have

$$\mathrm{Log}\lambda_n(XY^{-1}) \le \delta(X,Y) \text{ and } \mathrm{Log}\lambda_1(XY^{-1}) \ge -\delta(X,Y).$$

Therefore, $2^{-\delta(X,Y)}I \le Y^{-1/2}XY^{-1/2} \le 2^{\delta(X,Y)}I$, which completes the proof. □

Next, we consider a deterministic sequence $\{z_k\}_{k\in\mathbb{N}}$ with each z_k taking value from $\{0,1\}$. Associated with the sequence $\{z_k\}_{k\in\mathbb{N}}$ we define the (deterministic) recursion:

$$P_{k+1} = AP_kA' + Q - z_kAP_kC'(CP_kC' + R)^{-1}CP_kA'. \quad (13)$$

The following lemma holds.

Lemma 7: Consider the deterministic evolution (13).

(i) If there exists an initial condition $P_0 = \Sigma \in \mathbb{S}^n_+$ such that $\limsup_{k\to\infty} \mathrm{Tr}(P_k) = \infty$, then $\limsup_{k\to\infty} \mathrm{Tr}(P_k) = \infty$ for all $P_0 \in \mathbb{S}^n_+$;

(ii) If there exists an initial condition $P_0 = \Sigma \in \mathbb{S}^n_+$ such that $\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \infty$, then $\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \infty$ for all $P_0 \in \mathbb{S}^n_+$.

Proof: Consider two Kalman filters that undergo the packet loss process $\{z_k\}_{k\in\mathbb{N}}$: one has initial condition $\Sigma_1 \in \mathbb{S}^n_+$ while the other has initial condition $\Sigma_2 \in \mathbb{S}^n_+$. Denote the prediction error covariance matrices at time k from initial points Σ_1 and Σ_2, respectively, by $P_k^{\Sigma_1}$ and $P_k^{\Sigma_2}$. From Lemma 2, we can always find a sufficiently large integer t such that $g^t(\Sigma_1), g^t(\Sigma_2) \in \mathbb{S}^n_{++}$. According to (8) and (9), we have $P_t^{\Sigma_1}, P_t^{\Sigma_2} \in \mathbb{S}^n_{++}$. On the other hand, $P_t^{\Sigma_1} \le h^t(\Sigma_1)$ and $P_t^{\Sigma_2} \le h^t(\Sigma_2)$ by (9). Therefore, there always exists a constant $d \ge 0$ such that $\delta(P_t^{\Sigma_1}, P_t^{\Sigma_2}) \le d$. The fact from Lemma 5 that the operators h and g are non–expansive in $(\mathbb{S}^n_{++}, \delta)$ provided that A is invertible leads to

$$\delta(P_k^{\Sigma_1}, P_k^{\Sigma_2}) \le d$$

for all $k \ge t$. By Lemma 6,

$$P_k^{\Sigma_1} \ge \beta P_k^{\Sigma_2} \quad (14)$$

holds for all $k \ge t$, where $\beta \triangleq 2^{-d}$.

The unboundness of $\mathrm{Tr}(P_k^{\Sigma})$ means that, for any positive number C, there always exists a sufficiently large integer $\mathrm{N} \ge t$ such that $\mathrm{Tr}(P_{\mathrm{N}}^{\Sigma}) > \mathrm{C}$. When taking $\Sigma_1 = 0$ and $\Sigma_2 = \Sigma$ in (14), we have $\mathrm{Tr}(P_{\mathrm{N}}^{0}) \ge \beta\mathrm{C}$. By (7) and (8) again, $\mathrm{Tr}(P_{\mathrm{N}}^{P_0}) \ge \mathrm{Tr}(P_{\mathrm{N}}^{0}) \ge \beta\mathrm{C}$ holds for any $P_0 \in \mathbb{S}^n_+$. Since C is arbitrarily chosen, the assertion follows as claimed. The same is true of the statement (ii) as the contraposition of (i). □

There corresponds an analogy for $\liminf_{k\to\infty} \mathrm{Tr}(P_k)$ as we will present below. We omit the proof since it is similar to the proof of Lemma 7.

Lemma 8: Consider the deterministic evolution (13).

(i) If there exists an initial condition $P_0 = \Sigma$ such that $\liminf_{k\to\infty} \mathrm{Tr}(P_k) = \infty$, then $\liminf_{k\to\infty} \mathrm{Tr}(P_k) = \infty$ for all $P_0 \in \mathbb{S}_+^n$;

(ii) If there exists an initial condition $P_0 = \Sigma$ such that $\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \infty$, then $\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \infty$ for all $P_0 \in \mathbb{S}_+^n$.

The definition of tail events is as follows:

Definition 5: Let $\{\xi_k\}_{k\in\mathbb{N}}$ be a sequence of random variables, and $\tilde{\mathcal{F}}_k \triangleq \sigma(\xi_k, \xi_{k+1}, \ldots)$ be the smallest σ–algebra generated by $\xi_k, \xi_{k+1}, \ldots$ Then, $\mathcal{T}(\{\xi_k\}_{k\in\mathbb{N}}) \triangleq \cap_{j=1}^{\infty} \tilde{\mathcal{F}}_j$ is called the tail algebra of $\{\xi_k\}_{k\in\mathbb{N}}$. If $\mathcal{A} \in \mathcal{T}(\{\xi_k\}_{k\in\mathbb{N}})$, then $\mathcal{A}$ is said to be a tail event of $\{\xi_k\}_{k\in\mathbb{N}}$.

We still need to recall the concept of *strong mixing*, which was first introduced in [34], and then the zero–one law for strong mixing random processes established in [35]. Note that $*$-mixing implies strong mixing [36].

Definition 6: The sequence of random variables $\{\xi_k\}_{k\in\mathbb{N}}$ on a probability space $(\mathscr{S}, \mathcal{S}, \mu)$ is said to be strong mixing if

$$\alpha(n) \triangleq \sup |\mu(\mathcal{A} \cap \mathcal{B}) - \mu(\mathcal{A})\mu(\mathcal{B})| \to 0, \quad \text{as } n \to \infty,$$

where the supremum is taken over all $\mathcal{A} \in \sigma(\xi_1, \ldots, \xi_k)$, $\mathcal{B} \in \sigma(\xi_{k+n}, \xi_{k+n+1}, \ldots)$ and $k \in \mathbb{N}$.

The following lemma holds.

Lemma 9 (Zero–One Law for Strong Mixing ([36, Th. 2.3])): Let $\{\xi_k\}_{k\in\mathbb{N}}$ be a sequence of strong mixing random variables on a probability space $(\mathscr{S}, \mathcal{S}, \mu)$. Let $\mathcal{T}(\{\xi_k\}_{k\in\mathbb{N}})$ be the tail algebra of $\{\xi_k\}_{k\in\mathbb{N}}$. Then for any $\mathcal{A} \in \mathcal{T}(\{\xi_k\}_{k\in\mathbb{N}})$, there holds $\mu(\mathcal{A}) = 1$ or 0.

Remark 1: The concepts of $*$-mixing and strong mixing introduced in Definitions 2 and 6, originated in [24] and [34], respectively, are imposed under different measures of dependence between past and future along the random sequence. The $*$-mixing is equipped with a tighter measure, and therefore implies the strong mixing. For a detailed introduction of their relations we hereby refer to [36], in which the $*$-mixing corresponds to ψ-mixing, and the strong mixing corresponds to α-mixing.

B. Proof of Theorem 1

We only focus on assertion (i) for the event $\mathcal{E} \triangleq \{\omega : \limsup_{k\to\infty} \mathrm{Tr}(P_k)(\omega) < \infty\}$, as the conclusion for (ii) can be proved using the same argument.

Since $\mathrm{Tr}(P_k)$ is $\mathcal{F}$–measurable for all $k \in \mathbb{N}$, so is $\limsup_{k\to\infty} \mathrm{Tr}(P_k)$. Then $\mathcal{E} \in \mathcal{F}$. Fix a positive integer n and a deterministic sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ with each $\tilde{z}_k$ taking value from $\{0, 1\}$. We define a sequence $\{z_k\}_{k\in\mathbb{N}}$, where $z_k \in \{0, 1\}$, such that $z_k = \tilde{z}_{k+n}$. Accordingly, we define a sequence of matrices $\{\tilde{P}_k\}_{k\in\mathbb{N}}$ as the estimation error covariances of the Kalman filter along $\{\tilde{z}_k\}_{k\in\mathbb{N}}$, where initial point is denoted by $\tilde{P}_0 \in \mathbb{S}_+^n$; and define $\{P_k\}_{k\in\mathbb{N}}$ along $\{z_k\}_{k\in\mathbb{N}}$, where initial point is denoted by $P_0 \in \mathbb{S}_+^n$. The rest of the proof consists of two aspects:

(a) Suppose that $\limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_k) = \infty$ holds with a given initial covariance $\tilde{P}_0$. When $P_0 = \tilde{P}_n$, we have $\limsup_{k\to\infty} \mathrm{Tr}(P_k) = \limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_{k+n}) = \infty$. It follows from (i) of Lemma 7 that $\limsup_{k\to\infty} \mathrm{Tr}(P_k) = \infty$ holds for any $P_0 \in \mathbb{S}_+^n$.

(b) Suppose that $\limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_k) < \infty$ holds along $\{\tilde{z}_k\}_{k\in\mathbb{N}}$. Then we have $\limsup_{k\to\infty} \mathrm{Tr}(P_k) = \limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_{k+n}) < \infty$ when $P_0 = \tilde{P}_n$. It follows from (ii) of Lemma 7 that $\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \infty$ holds for any $P_0 \in \mathbb{S}_+^n$.

Define

$$\mathbb{S}_0 \triangleq \{X : X = \phi_{\tilde{z}_n} \circ \cdots \circ \phi_{\tilde{z}_1}(\Sigma)\},$$

where $\tilde{z}_1, \ldots, \tilde{z}_n \in \{0, 1\}$, and ϕ_i equals to the mapping h when $i = 0$ and g when $i = 1$. It is straightforward that $\mathbb{S}_0$ is bounded (it is a finite set) and $\mathbb{S}_0 \subseteq \mathbb{S}_+^n$, therefore showing that whether $\limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_k) < \infty$ holds or not does not depend on $\tilde{z}_1, \ldots, \tilde{z}_n$. Since $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ is arbitrarily chosen from $\mathbf{\Omega}$, we conclude that the event $\mathcal{E}$ and its compliment $\mathcal{E}^c$ are independent of $\sigma(\gamma_1, \ldots, \gamma_n)$. Again since n is arbitrarily taken, $\mathcal{E} \in \mathcal{T}(\{\gamma_k\}_{k\in\mathbb{N}})$. The conclusion then follows from Lemma 9.

C. Proof of Theorem 2

First of all, we establish an auxiliary lemma.

Lemma 10: Suppose $\mathrm{I_o} = 1$. Then for any constant $\mathrm{C} > \mathrm{Tr}(M)$, where M is defined in (11), the events $\{\omega : \limsup_{k\to\infty} \mathrm{Tr}(P_k)(\omega) < \mathrm{C}\}$ and $\{\omega : \liminf_{k\to\infty} \mathrm{Tr}(P_k)(\omega) < \mathrm{C}\}$ are tail events of $\{\gamma_k\}_{k\in\mathbb{N}}$.

Proof: First let us show the conclusion for the event $\mathcal{A}_{\mathrm{C}} \triangleq \{\omega : \limsup_{k\to\infty} \mathrm{Tr}(P_k)(\omega) < \mathrm{C}\}$. As in the proof of Theorem 1, we can readily show $\mathcal{A}_{\mathrm{C}} \in \mathcal{F}$. Fix a positive integer n, a real number $\mathrm{C} > 0$ and a deterministic sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ with each $\tilde{z}_k$ taking value from $\{0, 1\}$. we define a sequence $\{z_k\}_{k\in\mathbb{N}}$, where $z_k \in \{0, 1\}$, such that $z_k = \tilde{z}_{k+n}$. Accordingly, define a sequence of matrices $\{\tilde{P}_k\}_{k\in\mathbb{N}}$ as the estimation error covariances along $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ with an initial point denoted by $\tilde{P}_0 \in \mathbb{S}_+^n$; and define $\{P_k\}_{k\in\mathbb{N}}$ along $\{z_k\}_{k\in\mathbb{N}}$ with an initial point denoted by $P_0 \in \mathbb{S}_+^n$. The rest of the proof consists of two aspects:

(a) Suppose that $\limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_k) < \mathrm{C}$ holds with a given initial point $\tilde{P}_0$. In light of Lemma 3, we conclude via reduction to absurdity that $\{\tilde{z}_k\}_{k\in\mathbb{N}} \in \{\omega : \omega \in \gamma_k = 1, i.o.\}$. When $P_0 = \tilde{P}_n \triangleq \Sigma_1$, we have $\limsup_{k\to\infty} \mathrm{Tr}(P_k) = \limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_{k+n}) < \mathrm{C}$. Next we shall now show $\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \mathrm{C}$ for any $P_0 \in \mathbb{S}_+^n$. Choose any matrix $\Sigma_2 \in \mathbb{S}_+^n$. We differentiate P_k with different initial points Σ_1 and Σ_2 by using notations $P_k^{\Sigma_1}$ and $P_k^{\Sigma_2}$ respectively. By Lemma 2, there exists an integer t such that $P_t^{\Sigma_1}, P_t^{\Sigma_2} \in \mathbb{S}_{++}^n$. Since $\mathrm{I_o} = 1$, Lemma 5 indicates that the operator g is strictly contractive in $(\mathbb{S}_{++}^n, \delta)$. Therefore,

$$\delta(P_k^{\Sigma_1}, P_k^{\Sigma_2}) \le q^{\sum_{i=t}^{k} z_i} \delta(P_t^{\Sigma_1}, P_t^{\Sigma_2})$$

holds for all $k \geq t$, where $q \in (0, 1)$ is a constant that only depends on A, C, Q, R. As $k \to \infty$, we have $\sum_{i=t}^{k} z_i \to \infty$ and consequently $\delta(P_k^{\Sigma_1}, P_k^{\Sigma_2}) \to 0$. Thus, $P_k^{\Sigma_2} \to P_k^{\Sigma_1}$ due to the fact that $(\mathbb{S}_{++}^n, \delta)$ is a complete metric space. Since C is arbitrarily chosen, $\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \mathrm{C}$ holds for any $P_0 \in \mathbb{S}_+^n$.

(b) On the other hand, we suppose that $\limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_k) \geq \mathrm{C}$ holds with a given initial covariance $\tilde{P}_0 \in \mathbb{S}_+^n$. We discuss in all cases: $\limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_k)$ is bounded by a larger constant $\tilde{\mathrm{C}} > \mathrm{C}$ or unbounded. For the first case, by using the same argument as in (i), $\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \tilde{\mathrm{C}}$ holds for any $P_0 \in \mathbb{S}_+^n$. For the other case, it follows form the proof of Theorem 1 that $\limsup_{k\to\infty} \mathrm{Tr}(P_k) = \infty$ holds for any $P_0 \in \mathbb{S}_+^n$.

Define

$$\mathbb{S}_0 \triangleq \left\{X : X = \phi_{\tilde{z}_n} \circ \cdots \circ \phi_{\tilde{z}_1}(\Sigma)\right\},$$

where $\tilde{z}_1, \ldots, \tilde{z}_n \in \{0, 1\}$, and ϕ_i equals to the mapping h when $i = 0$ and g when $i = 1$. It is straightforward that $\mathbb{S}_0$ is bounded (it is a finite set) and $\mathbb{S}_0 \subseteq \mathbb{S}_+^n$. In view of the arguments in (i) and (ii), we obtain that whether $\limsup_{k\to\infty} \mathrm{Tr}(\tilde{P}_k) < \mathrm{C}$ holds or not does not depend on $\tilde{z}_1, \ldots, \tilde{z}_n$. Since $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ is arbitrarily chosen, the event $\mathcal{A}_\mathrm{C}$ and its compliment $(\mathcal{A}_\mathrm{C})^c$ are independent of $\sigma(\gamma_1, \ldots, \gamma_n)$. Again since n and C are arbitrarily taken, $\mathcal{A}_\mathrm{C} \in \mathcal{T}(\{\gamma_k\}_{k\in\mathbb{N}})$ holds for all $\mathrm{C} > 0$.

It remains to show the assertion for the event $\mathcal{E}_\mathrm{C} \triangleq \{\omega : \liminf_{k\to\infty} \mathrm{Tr}(P_k)(\omega) < \mathrm{C}\}$. On one hand, by reduction to absurdity it is true for any $\mathrm{C} > \mathrm{Tr}(M)$ that

$$\mathcal{E}_\mathrm{C} \subseteq \{\omega : \omega \in \gamma_k = 1 \ i.o.\}.$$

On the other hand, from Lemma 4,

$$\begin{aligned}\{\omega : \omega \in \gamma_k = 1, i.o.\} &\subseteq \{\omega : \liminf_{k\to\infty} \mathrm{Tr}(P_k)(\omega) < \mathrm{Tr}(M)\}\\ &\subseteq \mathcal{E}_\mathrm{C},\end{aligned} \tag{15}$$

where the second "$\subseteq$" holds since $\mathrm{C} > \mathrm{Tr}(M)$. In summary, we have

$$\mathcal{E}_\mathrm{C} = \{\omega : \omega \in \gamma_k = 1 \ i.o.\}. \tag{16}$$

Then $\mathcal{E}_\mathrm{C} \in \mathcal{T}(\{\gamma_k\}_{k\in\mathbb{N}})$ as one realizes that the latter event in (16) is a tail event. □

We are now in a place to complete the proof of Theorem 2. We only focus on the statement for absolutely upper a.s. stability, since that for absolutely lower a.s. stability can be analogously proved. Define

$$\mathcal{A}_x \triangleq \{\omega : \limsup_{k\to\infty} \mathrm{Tr}(P_k)(\omega) < x\}, \quad x > 0.$$

It is clear that $\mathcal{A}_{\lfloor x \rfloor} \subseteq \mathcal{A}_x \subseteq \mathcal{A}_{\lceil x \rceil}$ holds for all $x > 0$, which eventually results in

$$\bigcup_{\mathrm{C}\in(0,\infty)} \mathcal{A}_\mathrm{C} = \bigcup_{\mathrm{C}\in\mathbb{N}} \mathcal{A}_\mathrm{C}.$$

Since $\mathcal{A}_\mathrm{C} \in \mathcal{T}(\{\gamma_k\}_{k\in\mathbb{N}})$ for any $\mathrm{C} > \mathrm{Tr}(M)$ by Lemma 10,

$$\bigcup_{\mathrm{C}\in(0,\infty)} \mathcal{A}_\mathrm{C} = \bigcup_{\substack{\mathrm{C}\in\mathbb{N},\\ \mathrm{C}>\mathrm{Tr}(M)}} \mathcal{A}_\mathrm{C} \in \mathcal{T}(\{\gamma_k\}_{k\in\mathbb{N}}).$$

Finally, the conclusion follows from Lemma 9.

IV. Almost Sure Stability Conditions

In the last section, we have shown that whether the considered Kalman filter is a.s. stable or not can be interpreted by a zero–one law. In this section, we are devoted to studying the relationship between the packet rate and these stability notions. We first present some sufficient/necessary stability conditions for general LTI systems. Then we continue to show that one–step observable systems admit tighter results, with necessary and sufficient conditions derived for upper and lower a.s. stabilities, respectively. Finally, for the so–called non–degenerate systems, we give a necessary and sufficient upper a.s. stability condition.

Denote $\mathbb{E}[\gamma_k] \triangleq p_k$. To make the analysis concise, we require the following assumption

(A3) $\{p_k\}_{k\in\mathbb{N}}$ *is a monotonic sequence.*

It is not difficult to find that, if (A3) is not satisfied, all results are still tractable under the current analysis but in more complex forms. We choose (A3) to be our standing assumption in the rest of this section.

A. Main Results

1) General Stability Conditions: First we give sufficient conditions for (absolute) lower a.s. stability and lower a.s. instability. Recall that $\mathrm{I_o}$ is the observability index defined in Definition 1.

Theorem 3: Let Assumptions (A1)–(A3) hold.

(i) If $\sum_{k=1}^{\infty} (p_k)^{\mathrm{I_o}} = \infty$, then the considered filtering system is absolutely lower a.s. stable for any $P_0 \in \mathbb{S}_+^n$.

(ii) If $\sum_{k=1}^{\infty} p_k < \infty$, then the considered filtering system is lower a.s. unstable for any $P_0 \in \mathbb{S}_+^n$.

The following theorem presents a necessary condition for upper a.s. stability.

Theorem 4: Let Assumptions (A1)–(A3) hold. If the considering filtering system is upper a.s. stable, then there exists a constant $\mathrm{I} \in \mathbb{N}$ such that $\sum_{k=1}^{\infty} (1 - p_k)^{\mathrm{I}} < \infty$.

The proofs of Theorems 3 and 4 rely on Borel-Cantelli lemmas with respect to $*$-mixing.

2) One–Step Observable Systems: As a special case, one–step observable systems have $\mathrm{I_o} = 1$. The following theorem provides necessary and sufficient conditions for (absolutely) lower a.s. stability.

Theorem 5: Let Assumptions (A1)–(A3) hold. Suppose $\mathrm{I_o} = 1$. For any $P_0 \in \mathbb{S}_+^n$, the following conditions are equivalent:

(i) The considered filtering system is absolutely lower a.s. stable;

(ii) The considered filtering system is lower a.s. stable;

(iii) There holds that $\sum_{k=1}^{\infty} p_k = \infty$.

In the following, we also present necessary and sufficient conditions for (absolutely) upper a.s. stability.

Theorem 6: Let Assumptions (A1)–(A3) hold. Suppose $\mathrm{I_o} = 1$. For any $P_0 \in \mathbb{S}_+^n$, the following statements are equivalent:

(i) The considered filtering system is absolutely upper a.s. stable;
(ii) The considered filtering system is upper a.s. stable;
(iii) There exists a constant $\mathrm{I} \in \mathbb{N}$ such that $\sum_{k=1}^{\infty}(1-p_k)^{\mathrm{I}} < \infty$.

Theorems 5 and 6 are proved, partially relying on the fact that, as long as the considered Kalman filter successfully receives a packet, its instantaneous error covariance is bounded from above. The detailed proofs have been put in Section IV-C.

3) Non–Degenerate Systems: For general LTI systems with $\mathrm{I_o} \geq 2$, it is challenging to find conditions guaranteeing (absolutely) upper a.s. stability, since P_k does not necessarily decrease when packets are intermittently received. However, an exception is a class of so–called non–degenerate systems. We first introduce the definition of non–degenerate systems, which is taken from [14] and [15], and then present the probabilistic stability guarantor of $\sup_{k\geq n} \mathrm{Tr}(P_k)$ for this kind of systems. Note that the requirement of non–degeneracy is indispensable because it enables us to bound $\mathrm{Tr}(P_k)$ when intermittent receptions of measurements happen.

Definition 7: Consider a system (C, A) in diagonal standard form, i.e., $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $C = [C_1, \ldots, C_n]$. A quasi–equiblock of the system is defined as a subsystem $(C_{\mathcal{I}}, A_{\mathcal{I}})$, where $\mathcal{I} \triangleq \{l_1, \ldots, l_i\} \subset \{1\ldots, n\}$, such that $A_{\mathcal{I}} = \mathrm{diag}(\lambda_{l_1}, \ldots, \lambda_{l_i})$ with $|\lambda_{l_1}| = \cdots = |\lambda_{l_i}|$ and $C_{\mathcal{I}} = [C_{l_1}, \ldots, C_{l_i}]$.

Definition 8: A diagonalizable system (C, A) is non–degenerate if every quasi–equiblock of the system is one–step observable. Conversely, it is degenerate if it has at least one quasi–equiblock that is not one–step observable.

The following result holds.

Theorem 7: Let Assumptions (A1)–(A3) hold. Suppose the system (C, A) is non–degenerate. For any $P_0 \in \mathbb{S}_+^n$, the following statements are equivalent:

(i) The considered filtering system is absolutely upper a.s. stable;
(ii) The considered filtering system is upper a.s. stable;
(iii) There exists a constant $\mathrm{I} \in \mathbb{N}$ such that $\sum_{k=1}^{\infty}(1-p_k)^{\mathrm{I}} < \infty$.

Remark 2: The necessary and sufficient conditions in Theorem 7 suggest that, when (C, A) is non–degenerate, the absolutely upper a.s. stability also follows a zero–one law.

B. Supporting Lemmas

This subsection presents supporting lemmas and auxiliary definitions for the proofs of the main results. The following two lemmas concern with sequences of real numbers. The first one is well known and its proof can be found in [37].

Lemma 11: Suppose that $\{a_k\}_{k\in\mathbb{N}}$ is a sequence of real numbers with $a_k \in [0, 1)$. Then $\sum_{k=1}^{\infty} a_k = \infty$ holds if and only if $\prod_{k=1}^{\infty}(1 - a_k) = 0$.

Lemma 12: Suppose that $\{a_k\}_{k\in\mathbb{N}}$ is a monotonic sequence of real numbers with $a_k \in [0, \infty)$. Then, for any $l \geq 2$, $\sum_{i=0}^{\infty} \prod_{k=il+1}^{(i+1)l} a_k = \infty$ holds if and only if $\sum_{k=1}^{\infty}(a_k)^l = \infty$.

Proof: Without loss of generality, we assume that $\{a_k\}_{k\in\mathbb{N}}$ is monotonically decreasing, for a monotonically increasing sequence can be treated in a similar manner. For simplicity, let $s_j \triangleq \sum_{i=1}^{\infty} \prod_{k=(i-1)n+j}^{in+j-1} a_k$ for $j \in \mathbb{N}$. If $s_1 = \infty$, observing that $s_1 \geq s_2 \geq \cdots \geq s_n \geq s_{n+1}$, and that $s_{n+1} = s_1 - \prod_{k=1}^{n} a_k$, we have $s_j = \infty$. Therefore,

$$\sum_{j=1}^{n} s_j \leq \sum_{k=1}^{\infty}(a_k)^n,$$

implying $\sum_{k=1}^{\infty}(a_k)^n = \infty$. To prove the sufficiency, note that

$$ns_1 \geq \sum_{j=1}^{n} s_j \geq \sum_{k=n}^{\infty}(a_k)^n.$$

Since n is finite, the desired conclusion follows. □

The following lemma is the first Borel–Cantelli lemma from probability theory. For more details, please refer to [29].

Lemma 13 (First Borel–Cantelli Lemma): Let $(\mathscr{S}, \mathcal{S}, \mu)$ be a probability space. Suppose $\{\mathcal{A}_j\}_{j\in\mathbb{N}}$ is a sequence of events, where $\mathcal{A}_j \in \mathcal{S}$ for all $j \in \mathbb{N}$. If $\sum_{i=1}^{\infty} \mu(\mathcal{A}_i) < \infty$, then $\mu(\mathcal{A}_i \ i.o.) = 0$.

The definition of $*$-mixing for a sequence of events, and the corresponding second Borel–Cantelli lemma are as follows.

Definition 9: A sequence of events $\{\mathcal{A}_j\}_{j\in\mathbb{N}}$ is said to be $*$-mixing if $\{1_{\mathcal{A}_j}\}_{j\in\mathbb{N}}$ is $*$-mixing.

Lemma 14 (Second Borel–Cantelli Lemma Under $*$–Mixing ([24, Lemma 6]): Let $\{\mathcal{A}_j\}_{j\in\mathbb{N}}$ be a sequence of $*$-mixing events on a probability space $(\mathscr{S}, \mathcal{S}, \mu)$. Then $\mu(\mathcal{A}_i \ i.o.) = 1$ if $\sum_{i=1}^{\infty} \mu(\mathcal{A}_i) = \infty$.

We define the following two quantities to evaluate the minimum and maximum lengths of consecutive packet drops that make the error covariance exceed a given threshold. With the help of the two quantities, we develop a sufficient condition for that $\limsup_{k\to\infty} \mathrm{Tr}(P_k)$ exceeds a given threshold almost surely. For a one–step observable system (i.e., $\mathrm{I_o} = 1$), as long as one packet is received, $P_k \leq M$ holds by Lemma 4, enabling us to develop necessary conditions for such a system.

Let us define two quantities $\overline{\mathrm{I}}(\mathrm{C})$ and $\underline{\mathrm{I}}(\mathrm{C})$ as follow: for a given real number $\mathrm{C} \geq \mathrm{Tr}(M)$, put

$$\overline{\mathrm{I}}(\mathrm{C}) \triangleq \min\left\{k \in \mathbb{N} : \mathrm{Tr}\big(h^k(M)\big) > \mathrm{C}\right\}, \tag{17}$$

$$\underline{\mathrm{I}}(\mathrm{C}) \triangleq \min\left\{k \in \mathbb{N} : \mathrm{Tr}\big(h^k(\overline{P})\big) > \mathrm{C}\right\}. \tag{18}$$

Similar definitions for $\overline{\mathrm{I}}(\mathrm{C})$ and $\underline{\mathrm{I}}(\mathrm{C})$ primarily appeared in [10], where the quantities were used to derive upper and lower bounds on $\mathbb{P}\left(P_{k|k} \leq P_*\right)$ for some $P_* \in \mathbb{S}_+^n$. Different from [10], in this paper, we will use these two quantities to characterize the relationships between the packet rate and various stability notations in Definition 3.

The following lemma says that, for any $\mathrm{C} \geq \mathrm{Tr}(M)$, both $\bar{\mathrm{I}}(\mathrm{C})$ and $\underline{\mathrm{I}}(\mathrm{C})$ are bounded.

Lemma 15: Suppose A is unstable. Then, there holds $\bar{\mathrm{I}}(\mathrm{C}) \leq \underline{\mathrm{I}}(\mathrm{C}) < \infty$ for all $\mathrm{C} \geq \mathrm{Tr}(M)$.

Proof: First of all, it is evident from Lemma 1 that $\bar{\mathrm{I}}(\mathrm{C}) \leq \underline{\mathrm{I}}(\mathrm{C})$. Since $\overline{P} < M$ by Lemma 4, to show that $\bar{\mathrm{I}}(\mathrm{C})$ and $\underline{\mathrm{I}}(\mathrm{C})$ are finite for any $\mathrm{C} \geq \mathrm{Tr}(M)$, it suffices to show that there exists an integer $k \in \mathbb{N}$ implying $\mathrm{Tr}\left(h^k(\overline{P})\right) > \mathrm{C}$. By Lemma 3, there always exists an $a > 0$ such that $\mathrm{Tr}\left(h^k(X)\right) \geq a|\lambda_1(A)|^{2k}$. Therefore, when taking

$$k \geq \left\lceil \frac{\mathrm{Log}\mathrm{C} - \mathrm{Log}a}{2\mathrm{Log}|\lambda_1(A)|} \right\rceil + 1,$$

we have $a|\lambda_1(A)|^{2k} > \mathrm{C}$, which completes the proof. $\square$

Lemma 16: Suppose that $\mathrm{I_o} = 1$. Consider a real number $\mathrm{C} \geq \mathrm{Tr}(M)$. If $\sum_{k=1}^{\infty} (1-p_k)^{\bar{\mathrm{I}}(\mathrm{C})} < \infty$, then $\mathbb{P}\big(\limsup_{k\to\infty} \mathrm{Tr}(P_k) \leq \mathrm{C}\big) = 1$ holds for all $P_0 \in \mathbb{S}_+^n$.

Proof: Noticing

$$\prod_{k=i}^{\bar{\mathrm{I}}(\mathrm{C})+i-1} (1-p_k) \leq \max\left\{(1-p_k)^{\bar{\mathrm{I}}(\mathrm{C})} : i \leq k \leq \bar{\mathrm{I}}(\mathrm{C})+i-1\right\} \leq \sum_{k=i}^{\bar{\mathrm{I}}(\mathrm{C})+i-1} (1-p_k)^{\bar{\mathrm{I}}(\mathrm{C})},$$

we have

$$\sum_{i=1}^{\infty} \prod_{k=i}^{\bar{\mathrm{I}}(\mathrm{C})+i-1} (1-p_k) \leq \sum_{i=1}^{\infty} \sum_{k=i}^{\bar{\mathrm{I}}(\mathrm{C})+i-1} (1-p_k)^{\bar{\mathrm{I}}(\mathrm{C})}. \quad (19)$$

Since $\bar{\mathrm{I}}(\mathrm{C})$ is a finite number, each term $(1-p_k)^{\bar{\mathrm{I}}(\mathrm{C})}$ appears at most $\bar{\mathrm{I}}(\mathrm{C})$ times in the summation of the right-hand side of (19). As a result,

$$\sum_{i=1}^{\infty} \sum_{k=i}^{\bar{\mathrm{I}}(\mathrm{C})+i-1} (1-p_k)^{\bar{\mathrm{I}}(\mathrm{C})} \leq \bar{\mathrm{I}}(\mathrm{C}) \sum_{k=1}^{\infty} (1-p_k)^{\bar{\mathrm{I}}(\mathrm{C})}.$$

This leads to

$$\sum_{i=1}^{\infty} \prod_{k=i}^{\bar{\mathrm{I}}(\mathrm{C})+i-1} (1-p_k) \leq \bar{\mathrm{I}}(\mathrm{C}) \sum_{k=1}^{\infty} (1-p_k)^{\bar{\mathrm{I}}(\mathrm{C})} < \infty,$$

from the standing hypothesis.

Next, from Lemma 13 and the definition of $\bar{\mathrm{I}}(\mathrm{C})$, we see $\mathbb{P}\left(\mathrm{Tr}(P_k) > \mathrm{C}\ i.o.\right) = 0$ for any $P_0 \in \mathbb{S}_+^n$ since the event $\mathrm{Tr}(P_k) > \mathrm{C}$ requires that at least $\bar{\mathrm{I}}(\mathrm{C})$ consecutive drops happen before that. This completes the proof. $\square$

C. Proofs of Statements

1) Proof of Theorem 3: If $\sum_{k=1}^{\infty} (p_k)^{\mathrm{I_o}} = \infty$, by Lemma 12, one obtains $\sum_{i=0}^{\infty} \prod_{k=i\mathrm{I_o}+1}^{(i+1)\mathrm{I_o}} p_k = \infty$.

Define

$$\mathcal{A}_j \triangleq \left\{\omega : \prod_{l=j\mathrm{I_o}+1}^{(j+1)\mathrm{I_o}} \gamma_l(\omega) = 1, \omega \in \Omega\right\}, \quad j \in \mathbb{N}.$$

Since $\{\gamma_k\}_{k\in\mathbb{N}}$ is $*$-mixing and $\mathrm{I_o} \leq n$, the sequence $\{\mathcal{A}_j\}_{j\in\mathbb{N}}$ of events induced by $\{\gamma_k\}_{k\in\mathbb{N}}$ is $*$-mixing by definition. By Lemmas 4 and 14, we have

$$\mathbb{P}\left(P_k \leq M\ i.o.\right) \geq \mathbb{P}(\mathcal{A}_j\ i.o.) = 1,$$

where the first assertion follows.

If $\sum_{k=1}^{\infty} p_k < \infty$, $\mathbb{P}\left(\gamma_k = 1\ i.o.\right) = 0$ holds by Lemma 13. Then, by Lemma 3, there holds

$$\mathbb{P}\Big(\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \infty\Big) = 0,$$

which completes the proof.

2) Proof of Theorem 4: We shall prove the contraposition of the theorem, viz. that, if $\sum_{k=1}^{\infty} (1-p_k)^{\mathrm{I}} = \infty$ for any $\mathrm{I} \in \mathbb{N}$, then the considered filtering system is upper a.s. unstable. To this end, fix any constant $\mathrm{C} \geq \mathrm{Tr}(M)$ and a realization $\omega \in \Omega$ of $\{\gamma_k\}_{k\in\mathbb{N}}$. By the definition of $\underline{\mathrm{I}}(\mathrm{C})$ in (18) and Lemma 15, we have $\mathrm{Tr}\left(h^{\underline{\mathrm{I}}(\mathrm{C})}(\overline{P})\right) > \mathrm{C}$ and $\underline{\mathrm{I}}(\mathrm{C}) < \infty$. Then, from the continuity of the matrix trace and h operators, there always exists a constant $\beta \in (0,1)$ such that $\mathrm{Tr}\left(h^{\underline{\mathrm{I}}(\mathrm{C})}(\beta\overline{P})\right) > \mathrm{C}$. Since $\lim_{k\to\infty} g^k(0) = \overline{P}$, there exists a sufficiently large $\mathrm{N}(\beta)$ that implies $g^k(0) \geq \beta\overline{P}$ for all $k \geq \mathrm{N}(\beta)$, see the proof of Lemma 2 at this statement. By (9) in Lemma 1, $P_k(\omega) \geq g^k(0) > \beta\overline{P}$ holds for all $k \geq \mathrm{N}(\beta)$. This observation therefore leads to $\mathrm{Tr}\big(h^{\underline{\mathrm{I}}(\mathrm{C})}(P_k(\omega))\big) > \mathrm{C}$ for all $k \geq \mathrm{N}(\beta)$. When taking all ω's within Ω into account, we have

$$\mathcal{E}_{\underline{\mathrm{I}}(\mathrm{C})} \subseteq (\mathcal{A}_\mathrm{C})^c, \quad (20)$$

where $\mathcal{E}_{\underline{\mathrm{I}}(\mathrm{C})} \triangleq \{\omega : \underline{\mathrm{I}}(\mathrm{C})$ numbers of consecutive packet dropouts occur i.o.$\}$ and $\mathcal{A}_\mathrm{C} \triangleq \{\omega : \limsup_{k\to\infty} \mathrm{Tr}(P_k(\omega)) \leq \mathrm{C}\}$. In addition, the hypothesis $\sum_{k=1}^{\infty} (1-p_k)^{\underline{\mathrm{I}}(\mathrm{C})} = \infty$ implies

$$\sum_{i=0}^{\infty} \prod_{k=i\underline{\mathrm{I}}(\mathrm{C})+1}^{(i+1)\underline{\mathrm{I}}(\mathrm{C})} (1-p_k) = \infty$$

by Lemma 12. Define

$$\mathcal{B}_j = \left\{\omega : \prod_{l=j\underline{\mathrm{I}}(\mathrm{C})+1}^{(j+1)\underline{\mathrm{I}}(\mathrm{C})} \big(1-\gamma_l(\omega)\big) = 1, \omega \in \Omega\right\}, \quad j \in \mathbb{N}.$$

Since $\{\gamma_k\}_{k\in\mathbb{N}}$ is $*$-mixing and $\underline{\mathrm{I}}(\mathrm{C}) < \infty$, the events $\{\mathcal{B}_j\}_{j\in\mathbb{N}}$ induced by $\{\gamma_k\}_{k\in\mathbb{N}}$ is $*$-mixing by definition. By virtue of Lemma 14, it implies that

$$\mathbb{P}(\mathcal{E}_{\underline{\mathrm{I}}(\mathrm{C})}) \geq \mathbb{P}(\mathcal{B}_i\ i.o.) = 1. \quad (21)$$

Since C is arbitrarily chosen from the interval $[\mathrm{Tr}(M), \infty)$,

$$\begin{aligned}\{\omega : \limsup_{k\to\infty} \mathrm{Tr}(P_k(\omega)) < \infty\} &= \bigcup_{\mathrm{C}\in[\mathrm{Tr}(M),\infty)} \mathcal{A}_\mathrm{C} \\ &\subseteq \bigcup_{\mathrm{C}\in[\mathrm{Tr}(M),\infty)} (\mathcal{E}_{\underline{\mathrm{I}}(\mathrm{C})})^c \subseteq \bigcup_{\underline{\mathrm{I}}(\mathrm{C})=1}^{\infty} (\mathcal{E}_{\underline{\mathrm{I}}(\mathrm{C})})^c,\end{aligned}$$

where the first "$\subseteq$" is from (20) and the second one is due to $\underline{\mathrm{I}}(\mathrm{C}) < \infty$. As a result,

$$\begin{aligned}\mathbb{P}(\limsup_{k\to\infty} \mathrm{Tr}(P_k) < \infty) &\leq \mathbb{P}\Big(\bigcup_{\underline{\mathrm{I}}(\mathrm{C})=1}^{\infty} (\mathcal{E}_{\underline{\mathrm{I}}(\mathrm{C})})^c\Big) \\ &\leq \sum_{\underline{\mathrm{I}}(\mathrm{C})=1}^{\infty} \Big(1 - \mathbb{P}(\mathcal{E}_{\underline{\mathrm{I}}(\mathrm{C})})\Big) \\ &= 0,\end{aligned}$$

in which the second inequality is due to subadditivity of measure $\mathbb{P}$ and the last equality is due to (21). This completes the proof.

3) Proof of Theorem 5: $(i) \Rightarrow (ii)$ is true from the definition in its own right. Since $\mathrm{I}_o = 1$ and the fact that lower a.s. stability follows the one–zero law, $(ii) \Rightarrow (iii)$ and $(iii) \Rightarrow (i)$ hold by Theorem 3.

4) Proof of Theorem 6: Note that (i) implies (ii) by definition and (ii) implies (iii) by Theorem 4. It remains to show $(iii) \Rightarrow (i)$. Take a constant C such that $\mathrm{C} \geq \min\{\mathrm{Tr}(M), h^{\mathrm{I}}(M)\}$. By (17) and Lemma 15, we have $\mathrm{I} < \overline{\mathrm{I}}(\mathrm{C}) < \infty$. Then (iii) implies $\sum_{k=1}^{\infty}(1-p_k)^{\overline{\mathrm{I}}(\mathrm{C})} < \infty$. According to Lemma 16, $\mathbb{P}\big(\limsup_{k\to\infty} \mathrm{Tr}(P_k) \leq \mathrm{C}\big) = 1$ holds for all $P_0 \in \mathbb{S}_+^n$, which completes the proof.

5) Proof of Theorem 7: Similar to the proof of Theorem 6, we only need to show $(iii) \Rightarrow (i)$. To this end, we first define a sequence of stopping time $\{t_j\}_{j\in\mathbb{N}}$ as a sequence of packet arrival times as follows:

$$\begin{aligned} t_1 &\triangleq \min\{k : k \geq 1, \gamma_k = 1\}, \\ &\vdots \\ t_j &\triangleq \min\{k : k > t_{j-1}, \gamma_k = 1\}. \end{aligned}$$

If $\max\{j : t_j \leq k\} \geq n$, it means that the estimator has received no less than n packets up to time k. In this case, we define

$$\begin{aligned} \tau_{k,1} &\triangleq k - t_i \quad \text{where } i = \max\{j : t_j \leq k\}, \text{ and} \\ \tau_{k,j} &\triangleq t_{i-j+2} - t_{i-j+1}, \quad \text{for } 2 \leq j \leq n. \end{aligned}$$

To get the desired result, we need the following lemma.

Lemma 17: If $\max\{j : t_j \leq k\} \geq n$ and the system is non–degenerate, then the following inequality holds:

$$\mathrm{Tr}(P_{k+1}) \leq a_0 \prod_{j=1}^{n} (|\lambda_1(A)| + \epsilon)^{2\tau_{k,j}},$$

where a_0 is a constant independent of $\tau_{k,j}$ and ϵ can be arbitrarily small.

Proof: The result can be readily established from [15, Th. 4] and the fact that $|\lambda_1(A)| \geq \cdots \geq |\lambda_n(A)|$. $\square$

If there exists an $\mathrm{I} \in \mathbb{N}$ such that $\sum_{k=0}^{\infty}(1-p_k)^{\mathrm{I}} < \infty$, we can always find a sufficiently large positive number C_{I} satisfying $\mathrm{C}_{\mathrm{I}} > a_0(|\lambda_1(A)| + \epsilon)^{2(n+\mathrm{I}-2)}$ for a small $\epsilon > 0$. Given any time $k \geq n + \mathrm{I} - 1$, we compute

$$\begin{aligned} &\mathbb{P}\Big(\mathrm{Tr}(P_{k+1}) > \mathrm{C}_{\mathrm{I}}\Big) \\ &\leq \mathbb{P}\Big(\mathrm{Tr}(P_{k+1}) > a_0(|\lambda_1(A)| + \epsilon)^{2(n+\mathrm{I}-2)}\Big) \\ &\leq \mathbb{P}\Big(\text{less than } n \text{ packets received between} \\ &\qquad\qquad \text{time } k-n-\mathrm{I}+2 \text{ and } k\Big) \\ &\leq \sum_{j=0}^{n-1} \binom{n+\mathrm{I}-1}{j} \max\{p_{k-n-\mathrm{I}+2}, p_k\}^j \\ &\quad \times (1 - \min\{p_{k-n-\mathrm{I}+2}, p_k\})^{n+\mathrm{I}-j-1} \\ &\leq \sum_{j=0}^{n-1} \binom{n+\mathrm{I}-1}{j} (1 - \min\{p_{k-n-\mathrm{I}+2}, p_k\})^{\mathrm{I}} \\ &\leq \sum_{j=0}^{n-1} \binom{n+\mathrm{I}-1}{j} (1 - p_{k-n-\mathrm{I}+2})^{\mathrm{I}} \\ &\quad + \sum_{j=0}^{n-1} \binom{n+\mathrm{I}-1}{j} (1 - p_k)^{\mathrm{I}} \end{aligned} \tag{22}$$

where the second inequality holds due to Lemma 17 and the observation that $\sum_{j=1}^{n} \tau_{k,j} \leq n + \mathrm{I} - 2$ if and only if less than n packets are received between time $k-n-\mathrm{I}+2$ and k, the second last inequality is from the monotonicity of $\{p_k\}_{k\in\mathbb{N}}$, and $\binom{\cdot}{\cdot}$ denotes a combination number. Thus,

$$\begin{aligned} &\sum_{k=1}^{\infty} \mathbb{P}\left(\mathrm{Tr}(P_k) > \mathrm{C}_{\mathrm{I}}\right) \\ &= \sum_{k=1}^{n+\mathrm{I}-1} \mathbb{P}\left(\mathrm{Tr}(P_k) > \mathrm{C}_{\mathrm{I}}\right) + \sum_{k=n+\mathrm{I}}^{\infty} \mathbb{P}\left(\mathrm{Tr}(P_k) > \mathrm{C}_{\mathrm{I}}\right) \\ &\leq \sum_{k=1}^{n+\mathrm{I}-1} \mathbb{P}\left(\mathrm{Tr}(P_k) > \mathrm{C}_{\mathrm{I}}\right) + 2\sum_{j=0}^{n-1} \binom{n+\mathrm{I}-1}{j} \sum_{k=1}^{\infty} (1-p_k)^{\mathrm{I}} \\ &< \infty, \end{aligned}$$

where the first inequality follows from (22). By Lemma 13, $\mathbb{P}\left(\mathrm{Tr}(P_k) > \mathrm{C}_{\mathrm{I}} \ i.o.\right) = 0$ holds even for the set of events $\{\omega : \mathrm{Tr}(P_k(\omega)) > \mathrm{C}_{\mathrm{I}}\}_{k\in\mathbb{N}}$ that are not independent, which completes the proof.

V. Conclusions

We have studied the stability, from the probabilistic perspective, of Kalman filtering with random packet dropouts. The packet dropouts were modeled by a $*$-mixing model, whereby the occurrence of any two packet drops can be considered approximately "independent" as they are sufficiently far apart from each other. We defined (absolutely) upper and lower a.s. stabilities of the considered filtering systems. We established a zero–one law of upper and lower a.s. stabilities for general LTI systems, which makes the upper and lower a.s. instabilities meaningful definitions, and when the filtering system is one–step observable, we showed that the absolutely upper and lower a.s. stabilities can also be interpreted using a zero–one law. To answer the "zero or one" question, we presented stability conditions for general LTI systems. When the system is one–step observable, it was further shown that absolutely a.s. stability is equivalent to a.s. stability, both of which are guaranteed by a necessary and sufficient condition in terms of packet arrival rates. Finally, for the so–called non–degenerate systems, a necessary and sufficient upper a.s. stability condition was given.

REFERENCES

[1] J. Wu, G. Shi, and K. H. Johansson, "Probabilistic convergence of Kalman filtering with nonstationary intermittent observations," in *Proc. 53rd IEEE Conf. Decis. Control*, Los Angeles, CA, USA, Dec. 2014, pp. 3783–3788.

[2] M. Franceschetti and P. Minero, "Elements of information theory for networked control systems," in *Information and Control in Networks*. New York, NY, USA: Springer, 2014, pp. 3–37.

[3] T. S. Rappaport, *Wireless Communications: Principles and Practice*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[4] T. J. Lim and Y. Ma, "The Kalman filter as the optimal linear minimum mean-squared error multiuser CDMA detector," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2561–2566, Nov. 2000.

[5] Q. M. Chaudhari, E. Serpedin, and J. S. Kim, "Energy-efficient estimation of clock offset for inactive nodes in wireless sensor networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 582–596, Jan. 2010.

[6] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proc. IEEE*, vol. 95, no. 1, pp. 138–162, Jan. 2007.

[7] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[8] X. Liu and A. Goldsmith, "Kalman filtering with partial observation losses," in *Proc. 43rd IEEE Conf. Decis. Control*, vol. 4, Dec. 2004, pp. 4180–4186.

[9] K. Plarre and F. Bullo, "On Kalman filtering for detectable systems with intermittent observations," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 386–390, Feb. 2009.

[10] L. Shi, M. Epstein, and R. M. Murray, "Kalman filtering over a packet-dropping network: A probabilistic perspective," *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 594–604, Mar. 2010.

[11] M. Huang and S. Dey, "Stability of Kalman filtering with Markovian packet losses," *Automatica*, vol. 43, no. 4, pp. 598–607, 2007.

[12] L. Xie and L. Xie, "Stability of a random Riccati equation with Markovian binary switching," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1759–1764, Aug. 2008.

[13] N. Xiao, L. Xie, and M. Fu, "Kalman filtering over unreliable communication networks with bounded Markovian packet dropouts," *Int. J. Robust Nonlinear Control*, vol. 19, no. 16, pp. 1770–1786, 2009.

[14] K. You, M. Fu, and L. Xie, "Mean square stability for Kalman filtering with Markovian packet losses," *Automatica*, vol. 47, no. 12, pp. 2647–2657, Dec. 2011.

[15] Y. Mo and B. Sinopoli, "Kalman filtering with intermittent observations: Tail distribution and critical value," *IEEE Trans. Autom. Control*, vol. 57, no. 3, pp. 677–689, Mar. 2012.

[16] D. E. Quevedo, A. Ahlén, and K. H. Johansson, "State estimation over sensor networks with correlated wireless fading channels," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 581–593, Mar. 2013.

[17] S. Kar, B. Sinopoli, and J. M. F. Moura, "Kalman filtering with intermittent observations: Weak convergence to a stationary distribution," *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 405–420, Feb. 2012.

[18] A. Censi, "Kalman filtering with intermittent observations: Convergence for semi-Markov chains and an intrinsic performance measure," *IEEE Trans. Autom. Control*, vol. 56, no. 2, pp. 376–381, Feb. 2011.

[19] L. Xie, "Stochastic comparison, boundedness, weak convergence, and ergodicity of a random Riccati equation with Markovian binary switching," *SIAM J. Control Optim.*, vol. 50, no. 1, pp. 532–558, 2012.

[20] D. Li, S. Kar, J. M. F. Moura, H. V. Poor, and S. Cui, "Distributed Kalman filtering over massive data sets: Analysis through large deviations of random Riccati equations," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1351–1372, Mar. 2015.

[21] Q. Zhang and S. A. Kassam, "Hybrid ARQ with selective combining for fading channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 5, pp. 867–880, May 1999.

[22] M. Stojanovic, "Underwater acoustic communications: Design considerations on the physical layer," in *Proc. 5th Annu. Conf. Wireless Demand Netw. Syst. Services*, 2008, pp. 1–10.

[23] G. Dib, "Vehicle-to-vehicle channel simulation in a network simulator," M.S. thesis, Inf. Netw. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2009.

[24] J. R. Blum, D. L. Hanson, and L. H. Koopmans, "On the strong law of large numbers for a class of stochastic processes," *Probab. Theory Rel. Fields*, vol. 2, no. 1, pp. 1–11, 1963.

[25] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2035–2041, Nov. 1994.

[26] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sep. 2005.

[27] R. M. Gray, *Entropy and Information Theory*. New York, NY, USA: Springer, 2011.

[28] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[29] R. Durrett, *Probability: Theory and Examples*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[30] P. Lancaster and L. Rodman, *Algebraic Riccati Equations*. London, U.K.: Oxford Univ. Press, 1995.

[31] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Upper Saddle River, NJ, USA: Prentice-Hall, 1979.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[33] P. Bougerol, "Kalman filtering with random coefficients and contractions," *SIAM J. Control Optim.*, vol. 31, no. 4, pp. 942–959, Jul. 1993.

[34] M. Rosenblatt, "A central limit theorem and a strong mixing condition," *Proc. Nat. Acad. Sci. USA*, vol. 42, no. 1, pp. 43–47, 1956.

[35] T. Nemetz and G. Varga, "On the extension of limit theorems," *Sel. Transl. Math. Statist. Probab.*, vol. 12, pp. 59–66, 1973.

[36] R. C. Bradley, "Basic properties of strong mixing conditions. A survey and some open questions," *Probab. Surv.*, vol. 2, no. 2, pp. 107–144, 2005.

[37] W. Rudin, *Real and Complex Analysis*. New York, NY, USA: McGraw-Hill, 1987.

Junfeng Wu received the B.Eng. from the Department of Automatic Control, Zhejiang University, Hangzhou, China, in 2009 and the Ph.D. degree in Electrical and Computer Engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2013. From September to December 2013, he was a Research Associate in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology, Hong Kong. From January 2014 to June 2017, He was a Postdoctoral Researcher at ACCESS (Autonomic Complex Communication nEtworks, Signals and Systems) Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden. He is currently with the College of Control Science and Engineering, Zhejiang University, P. R. China, His research interests include networked control systems, state estimation, and wirelesssensor networks, multi-agent systems. He received the Guan Zhao-Zhi Best Paper Award at the 34th Chinese Control Conference in 2015. He was selected to the national "1000-Youth Talent Program" of China in 2016.

Guodong Shi received the B.Sc. degree in mathematics and applied mathematics from the School of Mathematics, Shandong University, Jinan, China in 2005, and the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China in 2010. From 2010 to 2014, he was a Postdoctoral Researcher at the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden. Since May 2014, he has been with the Research School of Engineering, The Australian National University, Canberra, ACT, Australia, where he is now a Senior Lecturer and Future Engineering Research Leadership Fellow. His research interests include distributed control systems, quantum networking and decisions, and social opinion dynamics.

Brian D. O. Anderson (M'66–SM'74–F'75–LF'07) was born in Sydney, Australia, and educated at Sydney University in mathematics and electrical engineering, with PhD in electrical engineering from Stanford University in 1966. He is an Emeritus Professor at the Australian National University (having retired as Distinguished Professor in 2016), Distinguished Professor at Hangzhou Dianzi University, and Distinguished Researcher in National ICT Australia. His awards include the IEEE Control Systems Award of 1997, the 2001 IEEE James H Mulligan, Jr Education Medal, and the Bode Prize of the IEEE Control System Society in 1992, as well as several IEEE and other best paper prizes. He is a Fellow of the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, the Royal Society, and a foreign member of the US National Academy of Engineering. He holds honorary doctorates from a number of universities, including Universit Catholique de Louvain, Belgium, and ETH, Zrich. He is a past president of the International Federation of Automatic Control and the Australian Academy of Science. His current research interests are in distributed control, social networks and econometric modelling.

Karl Henrik Johansson is Director of the Stockholm Strategic Research Area ICT The Next Generation and Professor at the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology. He received M.Sc. and Ph.D. degrees from Lund University. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems, cyber-physical systems, and applications in transportation, energy, and automation. He is a member of the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and the European Control Association Council. He has received several best paper awards and other distinctions. He has been awarded Distinguished Professor with the Swedish Research Council and Wallenberg Scholar. He has received the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is Fellow of the IEEE and the Royal Swedish Academy of Engineering Sciences, and he is IEEE Distinguished Lecturer.