Probabilistic Convergence of Kalman Filtering with Nonstationary Intermittent Observations

Junfeng Wu[♯], Guodong Shi[♯] and Karl Henrik Johansson[♯]

Abstract— In this paper, we consider state estimation using a Kalman filter of a linear time-invariant process with nonstationary intermittent observations caused by packet losses. The packet loss process is modeled as a sequence of independent, but not necessarily identical Bernoulli random variables. Under this model, we show how the probabilistic convergence of the trace of the prediction error covariance matrices, which is denoted as $\mathrm{Tr}(P_k)$, depends on the statistical property of the nonstationary packet loss process. A series of sufficient and/or necessary conditions for the convergence of $\sup_{k \geq n} \mathrm{Tr}(P_k)$ and $\inf_{k \geq n} \mathrm{Tr}(P_k)$ are derived. In particular, for one-step observable linear system, a sufficient and necessary condition for the convergence of $\inf_{k \geq n} \mathrm{Tr}(P_k)$ is provided.

I. INTRODUCTION

Last decade has witnessed an increasing attention on networked control systems (NCSs) from the control, communication and networking communities, thanks to a rapid development of micro-electronics, wireless communication, and information and networking technologies. NCSs have applications in a wide range of areas in space exploration, environmental monitoring, health care, intelligent building, smart transportation and power grid [1]–[3], due to many advantages, including reducing operational cost, allowing distributed sensing and information sharing among different nodes, etc. Nevertheless, new challenges have also been introduced at the expense of all the aforementioned advantages. For instance, in NCSs, the information, represented in binary bits and transmitted in packets as the basic units [4] over a digital communication channel, is used to stabilize one or more dynamical systems, but packet dropouts and random delays, which ubiquitously exist in wireless communication, may deteriorate the performance of the closed-loop system or even cause the system to be unstable.

State estimation, offering the control unit with estimates of the system state based on the system outputs observed, is adopted in many networked control applications and is typically implemented recursively using a Kalman filter. The problem of Kalman filtering over a packet-dropping network was presented and investigated in [5], where modeled the packet losses as a Bernoulli process and studied how the Kalman filter evolves when packet losses are taken into consideration and how packet losses effect the stability of Kalman filtering. They proved that there exists a certain critical arrival rate for the packet losses, below which, $\mathbb{E}[P_k]$,

the expected value of the prediction error covariance matrix, is unbounded. Upper and lower bounds for this critical rate were provided. The above result was extended to partial observation losses in [6]. *Mo and Sinopoli* [7] and *Plarre and Bullo* [8] investigated the critical value under some less constrictive conditions for a linear system. To characterize the temporal correlation nature of practical channels, the Gilbert-Elliott model can be used to describe time-homogeneous Markovian packet arrivals. *Huang and Dey* [9] considered the same problem under the effect of Markovian packet losses. They introduced the notion of peak covariance and studied the peak covariance stability. Some sufficient conditions for peak covariance stability were provided in [10]. In [11], *You et al.* proved that the stability of Kalman filtering at packet reception times is equivalent to the stability at sampling time (i.e., the mean square stability). For some certain classes of linear systems, necessary and sufficient conditions for the mean square stability were derived. Essentially, the boundness of the moments of P_k can be fully characterized by its probability distribution function. Some related efforts have been made from a probabilistic perspective. *Shi et al.* [12] considered the calculation of $\mathbb{P}(P_k \leq M)$ and derived upper and lower bounds thereof. *Mo and Sinopoli* [13] analyzed how the trace of the estimation error covariance matrix decays. For non-degenerate systems, the critical arrival rate can be exactly derived from the decay rate.

Variance of channel's characteristics over time, such as in indoor radio channels and underwater acoustic channels, is difficult to model by a stationary random process [14]–[16], which is yet an important factor needs to be considered when designing control systems over wireless communication channels. Nonetheless, both i.i.d. packet-dropping model and the Gilbert-Elliott model as appeared in the aforementioned papers, fail to capture this basic nonstationary nature of communication channels. To fulfil the theoretical gap, we mainly focus on the probabilistic convergence of Kalman filtering with intermittent observations over a nonstationary fading channel. Note that the assumption, either that the packet loss process is identically distributed, or that the packet loss process has a stationary distribution imposed by the Gilbert-Elliott model which describes time-homogeneous Markovian packet arrivals, is relaxed in the packet loss model we considered, which is more realistic and provides a better modeling of the nonstationary channel fading. In this paper, packet losses are modeled as a sequence of independent, but not necessarily identical Bernoulli random variables. Under this model, neither the deterministic modified Riccati recursion with i.i.d. packet losses nor the stationary property

♯: ACCESS Linnaeus Center, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden. Email: {junfengw, guodongs, kallej}@kth.se.

imposed by the Markovian packet drops is valid. Rather than stability in mean square sense, the probabilistic convergence of Kalman filtering will be investigated in the present work. A series of sufficient and/or necessary conditions for the convergence of $\sup_{k \geq n} \mathrm{Tr}(P_k)$ and $\inf_{k \geq n} \mathrm{Tr}(P_k)$ are derived. In particular, for one-step observable linear system, a sufficient and necessary condition for the convergence of $\inf_{k \geq n} \mathrm{Tr}(P_k)$ is provided.

The remainder of the paper is organized as follows. Section II presents the problem setup. Section III introduces the main result. Conclusions and future work are given in the end.

Notations: $\mathbb{N}$ ($\mathbb{N}_+$) is the set of nonnegative (positive) integers. $k \in \mathbb{N}$ is time index. For a real number x, $\lceil x \rceil$ denotes the smallest integer not less than x. $\mathbb{S}_+^n$ is the set of n by n symmetric positive semi-definite matrices. $\sigma(\cdot)$ is the σ-algebra generated by random variables. For a matrix X, $\lambda_i(X)$, $i = 1, \ldots, n$ represents the ith largest eigenvalue of X. Moreover, $\|X\|_2$ means the spectral norm of the matrix X.

II. PROBLEM SETUP

Consider a linear time-invariant (LTI) process:

$$
\begin{align}
x_{k+1} &= Ax_k + w_k, \tag{1} \\
y_k &= Cx_k + v_k, \tag{2}
\end{align}
$$

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 $\mathbb{E}[w_k w_j{'}] = \delta_{kj} Q$ ($Q \geq 0$), $\mathbb{E}[v_k v_j{'}] = \delta_{kj} R$ ($R > 0$), $\mathbb{E}[w_k v_j{'}] = 0 \ \forall j, k$. 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 $\Sigma_0 \geq 0$. We assume that the pair (C, A) is observable and $(A, \sqrt{Q})$ controllable. It can be seen that, by applying a similarity transformation, the unstable and stable modes of the LTI system are decoupled. An open-loop prediction of the stable mode ever has a bounded estimation error covariance, therefore, this mode does not play any key in the problem considered in the present work. Without loss of generality, we henceforth assume that all of the eigenvalues of A have magnitudes not less than 1.
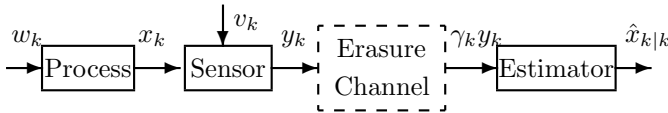

Fig. 1: Estimation over an erasure channel.

In this work, we consider an estimation scheme where the raw measurements of the sensor $\{y_k\}_{k \in \mathbb{N}_+}$ are transmitted to the estimator over an unreliable communication channel. The estimate, generated by the estimator, can be used as an input of control systems, or in monitoring applications. Although the leading motivation of the present work is networked systems with multi sensors and/or actuators pairwise connected

through a specific graph topology, the preliminary of understanding this problem is to consider a simple framework illustrated in Fig. 1. We denote by $\gamma_k \in \{0, 1\}$ the arrival of y_k at time k: If $\gamma_k = 1$, it indicates that y_k successfully arrives at the estimator; otherwise $\gamma_k = 0$. Motivated by nonstationary channel fading the communication network may suffer from, we assume that $\{\gamma_k\}_{k \in \mathbb{N}_+}$ is a sequence of independent but not necessarily identical random variables with $\mathbb{E}[\gamma_k] \triangleq \lambda_k \in (0, 1)$, and that $\{\gamma_k\}_{k \in \mathbb{N}_+}$ is also independent of how the system evolves. Although real digital communication introduces a bunch of other challenges, such as quantization and data rate, bit errors, and random delays, we are exclusively devoted to studying the impact of packet loss process on the estimation performance and therefore those effects will be ignored.

Define $\mathcal{F}_k$ as the filtration generated by all the measurements received by the estimator up to time k, i.e., $\mathcal{F}_k \triangleq \sigma(\gamma_t y_t, \gamma_t; 1 \leq t \leq k)$. 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 $\hat{x}_{k+1|k} = \mathbb{E}[x_{k+1} | \mathcal{F}_k]$. Let $P_{k|k}$ and $P_{k+1|k}$ be the corresponding estimation and prediction error covariance matrices, 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 [5]:

$$
\begin{align}
K_k &= P_{k|k-1} C' [C P_{k|k-1} C' + R]^{-1}, \tag{3} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + \gamma_k K_k (y_k - C\hat{x}_{k|k-1}), \tag{4} \\
P_{k|k} &= (I - \gamma_k K_k C) P_{k|k-1}, \tag{5} \\
\hat{x}_{k+1|k} &= A\hat{x}_{k|k}, \tag{6} \\
P_{k+1|k} &= A P_{k|k} A' + Q. \tag{7}
\end{align}
$$

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

$$
\begin{align}
h(X) &\triangleq AXA{'} + Q, \\
g(X) &\triangleq AXA' + Q - AXC{'}(CXC{'} + R)^{-1} CXA', \\
h^k(X) &\triangleq \underbrace{h \circ h \circ \cdots \circ h}_{k \text{ times}}(X) \quad \text{and} \quad g^k(X) \triangleq
\end{align}
$$

$\underbrace{g \circ g \circ \cdots \circ g}_{k \text{ times}}(X)$, where $\circ$ denotes the function composition. It is well known that, for a standard Kalman filter, $\lim_{k \to \infty} P_k = \overline{P}$ in which $\overline{P}$ is the unique positive semi-definite solution to $g(X) = X$ (see [17]).

III. MAIN RESULT

In this section, we will provide some results on the probabilistic convergence of Kalman filtering. Before we present the main result, let us first recall some preliminary lemmas.

Lemma 1: Suppose that $\{p_k\}_{k \in \mathbb{N}_+}$ is a monotonic sequence of real numbers with $p_k \in [0, \infty)$. Then, for any

$n \geq 2$, $\sum_{i=0}^{\infty} \prod_{k=in+1}^{(i+1)n} p_k = \infty$ holds if and only if $\sum_{k=1}^{\infty} (p_k)^n = \infty$.

Proof: Without loss of generality, assume that the sequence $\{p_k\}_{k\in\mathbb{N}_+}$ is monotonically decreasing. For a monotonically increasing sequence, the proof can be presented similarly. For simplicity, let $B_j \triangleq \sum_{i=1}^{\infty} \prod_{k=(i-1)n+j}^{in+j-1} p_k$ for $j \in \mathbb{N}_+$. If $B_1 = \infty$, observing that $B_1 \geq B_2 \geq \cdots \geq B_n \geq B_{n+1}$, and that $B_n = B_0 - \prod_{k=0}^{n-1} p_k$, we have $B_j = \infty$, $\forall\, j \in \mathbb{N}_+$. Therefore

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

which shows that $\sum_{k=1}^{\infty}(p_k)^n = \infty$. To prove the sufficient direction, note that

$$nB_1 \geq \sum_{j=1}^{n} B_j \geq \sum_{k=n}^{\infty}(p_k)^n.$$

Since n is finite, the conclusion follows. ∎

In what follows, in order to make the results below concise, we need to assume that $\{\lambda_k\}_{k\in\mathbb{N}_+}$ is a monotonic sequence in place. Without this assumption, it is not hard to verify that similar but complex results still hold.

Remark 1: Suppose that there exists an $\epsilon \in (0, 0.5)$ such that $\lambda_k \in [\epsilon, 1-\epsilon]$, $\forall\, k \in \mathbb{N}_+$. Then we can show that $\sum_{i=0}^{\infty} \prod_{k=in+1}^{(i+1)n} \lambda_k = \infty$ if and only if $\sum_{k=1}^{\infty}(\lambda_k)^n = \infty$ for any $n \geq 2$. To see this, observe that

$$\sum_{i=0}^{\infty} \prod_{k=i\,n+1}^{(i+1)n} \lambda_k \leq \left(\frac{1-\epsilon}{\epsilon}\right)^{n-1} \sum_{k=1}^{\infty}(\lambda_k)^n$$

and

$$\sum_{i=0}^{\infty} \prod_{k=i\,n+1}^{(i+1)n} \lambda_k \geq \left(\frac{\epsilon}{1-\epsilon}\right)^{n-1} \sum_{k=1}^{\infty}(\lambda_k)^n.$$

As $\epsilon \in (0, 0.5)$, the conclusion follows by the Squeeze Theorem. The assumption that $\lambda_k \in [\epsilon, 1-\epsilon]$, $\forall\, k \in \mathbb{N}_+$, where $\epsilon \in (0, 0.5)$, can replace the assumption of the monotonicity of $\{\lambda_k\}_{k\in\mathbb{N}_+}$ throughout the rest of this paper, without changing the conclusions.

The following two lemmas are well known, and their proofs are therefore omitted. For more details, please refer to [18] and [19], respectively.

Lemma 2: Suppose that $\{p_k\}_{k\in\mathbb{N}_+}$ is a sequence of real numbers with $p_k \in [0, 1)$. Then $\sum_{k=1}^{\infty} p_k = \infty$ holds if and only if $\prod_{k=1}^{\infty}(1-p_k) = 0$.

Lemma 3 (Borel-Cantelli Lemma): Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Assume that events $A_i \in \mathcal{F}$, $\forall\, i \in \mathbb{N}_+$. If $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$, then $\mathbb{P}(A_i, i.o.) = 0$, where "$A_i, i.o.$" means A_i occurs infinitely often. In addition, assuming that events A_i are independent, then $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ implies that $\mathbb{P}(A_i, i.o.) = 1$.

Before we proceed, let us introduce the definition of observability index which is first introduced in [9]. For an observable pair (A, C), define the *observability index* $\mathbb{I}_o \geq 1$

as the smallest integer such that $[C', A'C', \ldots, (A^{\mathbb{I}_o-1})'C']'$ is full-ranked. It is evident that $\mathbb{I}_o \leq n$. Define

$$M_0 \triangleq (J'J)^{-1}J' \left(H \begin{bmatrix} Q & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q \end{bmatrix} H' \right.$$
$$\left. + \begin{bmatrix} R & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R \end{bmatrix} \right) J(J'J)^{-1},$$

where $J \triangleq \left[(CA^{\mathbb{I}_o-1})', (CA^{\mathbb{I}_o-2})', \ldots, C' \right]'$ and

$$H = \begin{bmatrix} C & CA & \cdots & \cdots & CA^{\mathbb{I}_o-2} & 0 \\ 0 & C & \ddots & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & C & CA & \vdots \\ 0 & 0 & 0 & \cdots & C & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Denoted a constant matrix given as $\overline{M} = h^{\mathbb{I}_o-1}(M_0)$. For $\mathbb{I}_o$ and $\overline{M}$ defined above, we have the following lemma.

Lemma 4: If at time k there are at least $\mathbb{I}_o$ number of consecutive measurements $\{y_{k-\mathbb{I}_o+1}, \ldots, y_k\}$ received by the estimator, then $P_k \leq \overline{M}$.

Proof: Observe that

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-\mathbb{I}_o+1} \end{bmatrix} = Jx_{k-\mathbb{I}_o+1} + H \begin{bmatrix} w_{k-1} \\ w_{k-2} \\ \vdots \\ w_{k-\mathbb{I}_o+1} \end{bmatrix} + \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-\mathbb{I}_o+1} \end{bmatrix}.$$

Based on the consecutive measurements $\{y_{k-\mathbb{I}_o+1}, \ldots, y_k\}$ received by the estimator, one can use the following estimator to generate a linear estimate of x_k:

$$\tilde{x}_k = A^{\mathbb{I}_o-1}(J'J)^{-1}J' \begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-\mathbb{I}_o+1} \end{bmatrix}.$$

The associated estimation error covariance is exactly $\overline{M}$. Since Kalman filter is well known to be the optimal linear estimator, it implies that $P_k \leq \overline{M}$. ∎

Now we are ready to present the following theorem.

Theorem 1: Consider system (1) and (2) with $\mathbb{I}_o \geq 2$, and an independent packet loss process $\{\gamma_k\}_{k\in\mathbb{N}_+}$ of raw measurements with $\mathbb{E}[\gamma_k] = \lambda_k \in (0,1)$. Suppose that $\{\lambda_k\}_{k\in\mathbb{N}_+}$ is a monotonic sequence. For any $\Sigma_0 \geq 0$, if $\sum_{k=1}^{\infty}(\lambda_k)^{\mathbb{I}_o} = \infty$, then $\mathbb{P}\left(\liminf_{k\to\infty} \mathrm{Tr}(P_k) < \infty\right) = 1$ holds .

Proof: Lemma 4 implies that, at time k, $P_k \leq \overline{M}$ as long as $\{y_{k-\mathbb{I}_o+1}, \ldots, y_k\}$ is received by the estimator. Supposing that $\sum_{k=1}^{\infty}(\lambda_k)^{\mathbb{I}_o} = \infty$ and $\{\lambda_k\}_{k\in\mathbb{N}_+}$ is monotonic, one obtains from Lemma 1 that $\sum_{i=0}^{\infty} \prod_{k=i\mathbb{I}_o+1}^{(i+1)\mathbb{I}_o} \lambda_k = \infty$. By Lemma 3, it implies that $\mathbb{P}\left(P_k \leq \overline{M}, i.o.\right) = 1$. The proof is complete. ∎

As a special case, when C has a full column rank in (2), we have $\mathbb{I}_o = 1$. In this case, (A, C) is called to be one-step observable. And Theorem 1 is still valid even without the assumption that $\{\lambda_k\}_{k \in \mathbb{N}_+}$ is monotonic. Moreover, it becomes a necessary and sufficient condition. This result is presented as below.

Corollary 1: Consider system (1) and (2) with $\mathbb{I}_o = 1$, and an independent packet loss process $\{\gamma_k\}_{k \in \mathbb{N}_+}$ of raw measurements with $\mathbb{E}[\gamma_k] = \lambda_k \in (0, 1)$. For any $\Sigma_0 \geq 0$, if $\sum_{k=1}^{\infty} \lambda_k = \infty$, then $\mathbb{P}\left(\liminf_{k \to \infty} \mathrm{Tr}(P_k) < \infty\right) = 1$; if otherwise $\sum_{k=1}^{\infty} \lambda_k < \infty$, then $\mathbb{P}\left(\liminf_{k \to \infty} \mathrm{Tr}(P_k) < \infty\right) = 0$.

Proof: The proof of sufficiency directly follows from that of Theorem 1. For the necessity, we will show that if $\sum_{k=1}^{\infty} \lambda_k < \infty$ then $\mathbb{P}\left(\liminf_{k \to \infty} \mathrm{Tr}(P_k) < \infty\right) = 0$ in the following. In light of Lemma 3, the hypothesis $\sum_{k=1}^{\infty} \lambda_k < \infty$ implies $\mathbb{P}(\gamma_k = 1, i.o.) = 0$, which reveals that, for an $\omega \in \{0, 1\}^{\mathbb{N}_+}$, after a sufficient large $N(\omega)$, $\gamma_k = 0$, $k \geq N(\omega)$. Consequently, $P_k = h(P_{k-1})$, $\forall\, k \geq N(\omega)$. In light of Lemma 7, there exists no $M \in \mathbb{R}$ such that $\liminf_{k \to \infty} \mathrm{Tr}(P_k) \leq M$, i.e., $\mathbb{P}\left(\liminf_{k \to \infty} \mathrm{Tr}(P_k) < \infty\right) = 0$. ∎

In the sequel, we will consider the probabilistic convergence of $\sup_{k \geq n} \mathrm{Tr}(P_k)$. Let us first define two quantities – for a given $\overline{M} \geq \mathrm{Tr}(\overline{M})$, define $\overline{\mathbb{I}}(M)$ and $\underline{\mathbb{I}}(M)$ as follow:

$$\overline{\mathbb{I}}(M) = \min\{k \geq 1 : \mathrm{Tr}\left(h^k(\overline{M})\right) > M\}, \quad (8)$$
$$\underline{\mathbb{I}}(M) = \min\{k \geq 1 : \mathrm{Tr}\left(h^k(\overline{P})\right) > M\}. \quad (9)$$

Lemma 5: If A is unstable, for $\overline{\mathbb{I}}(M)$ and $\underline{\mathbb{I}}(M)$ defined in (8) and (9), it is true that

$$\overline{\mathbb{I}}(M) \leq \underline{\mathbb{I}}(M) < \infty, \ \forall\, M \geq \mathrm{Tr}(\overline{M}).$$

Proof: See Appendix. ∎

Similar definitions primarily appeared in [12]. In that paper, the two quantities are used to derive the upper and lower bounds of $\mathbb{P}\left(P_{k|k} \leq M\right)$, where M denotes a given positive semi-definite matrix. Different from [12], in this paper, we will use these two quantities to characterize the relationship between the boundness of $\limsup_{k \to \infty} \mathrm{Tr}(P_k)$ and the statistical property of $\{\gamma_k\}_{k \in \mathbb{N}_+}$. The following theorem provides with a counterpart of Theorem 1 for $\sup_{k \geq n} \mathrm{Tr}(P_k)$.

Theorem 2: Consider system (1) and (2), and an independent packet loss process $\{\gamma_k\}_{k \in \mathbb{N}_+}$ of raw measurements with $\mathbb{E}[\gamma_k] = \lambda_k \in (0, 1)$. Suppose that $\{\lambda_k\}_{k \in \mathbb{N}_+}$ is a monotonic sequence. Then, for a given positive real number $M \geq \mathrm{Tr}(\overline{M})$, $\mathbb{P}\left(\limsup_{k \to \infty} \mathrm{Tr}(P_k) > M\right) = 1$ holds for any $\Sigma_0 \geq 0$ if $\sum_{k=1}^{\infty}(1 - \lambda_k)^{\underline{\mathbb{I}}(M)} = \infty$.

Proof: From the definition in (9), we have that $\mathrm{Tr}\left(h^{\underline{\mathbb{I}}(M)}(\overline{P})\right) > M$. There must exist an $\epsilon \in (0, 1)$ such that $\mathrm{Tr}\left(h^{\underline{\mathbb{I}}(M)}(\overline{P} - \epsilon \overline{P})\right) > M$ still holds. Since $\lim_{k \to \infty} g^k(0) = \overline{P}$, there exists a sufficient large N and for any $t \geq N$ we have $0 \leq \overline{P} - g^t(0) \leq \epsilon \overline{P}$, i.e.,

$g^t(0) \geq (1 - \epsilon)\overline{P}$. Note that, for any $t \in \mathbb{N}_+$, it it evident that $P_t \geq g^t(0)$ because $\Sigma_0 \geq 0$ and $g(\cdot)$ is monotonically increasing. All the above observations lead to the fact that, for any $t \geq N$, $\mathrm{Tr}\left(h^{\underline{\mathbb{I}}(M)}(P_t)\right) > M$, i.e., $\mathrm{Tr}\left(h^{\underline{\mathbb{I}}(M)}(P_k)\right) > M$ always holds except for finite number of times.

On the other hand, by Lemma 1, the hypothesis implies that $\sum_{i=0}^{\infty} \prod_{k=i\underline{\mathbb{I}}(M)+1}^{(i+1)\underline{\mathbb{I}}(M)}(1 - \lambda_k) = \infty$ holds. Combining all the above observations, the assertion follows from Lemma 3. ∎

For some special systems, such as the system (1) is first-order or C has a full column rank, a necessary condition for $\mathbb{P}\left(\limsup_{k \to \infty} \mathrm{Tr}(P_k) > M\right) = 1$ can also be provided.

Theorem 3: Consider system (1) and (2) with $\mathbb{I}_o = 1$, and an independent packet loss process $\{\gamma_k\}_{k \in \mathbb{N}_+}$ of raw measurements with $\mathbb{E}[\gamma_k] = \lambda_k \in (0, 1)$. Then, for a given $M \geq \overline{M}$, $\mathbb{P}\left(\limsup_{k \to \infty} \mathrm{Tr}(P_k) > M\right) = 1$ holds only if $\sum_{k=1}^{\infty}(1 - \lambda_k)^{\overline{\mathbb{I}}(M)} = \infty$.

Proof: If $\sum_{k=1}^{\infty}(1 - \lambda_k)^{\overline{\mathbb{I}}(M)} < \infty$, one has that

$$\sum_{i=1}^{\infty} \prod_{k=i}^{\overline{\mathbb{I}}(M)+i-1}(1 - \lambda_k) \leq \sum_{i=1}^{\infty} \sum_{k=i}^{\overline{\mathbb{I}}(M)+i-1}(1 - \lambda_k)^{\overline{\mathbb{I}}(M)}$$
$$\leq \overline{\mathbb{I}}(M) \sum_{k=1}^{\infty}(1 - \lambda_k)^{\overline{\mathbb{I}}(M)} < \infty,$$

where the first inequality is from that $\overline{\mathbb{I}}(M) < \infty$, and that

$$\prod_{k=i}^{\overline{\mathbb{I}}(M)+i-1}(1 - \lambda_k) \leq \max\left\{(1 - \lambda_k)^{\overline{\mathbb{I}}(M)} : i \leq k \leq \overline{\mathbb{I}}(M) + i - 1\right\}$$
$$\leq \sum_{k=i}^{\overline{\mathbb{I}}(M)+i-1}(1 - \lambda_k)^{\overline{\mathbb{I}}(M)}.$$

Therefore it is straightforward to see that $\mathbb{P}\left(\mathrm{Tr}(P_k) \geq h^{\overline{\mathbb{I}}(M)}(\overline{M}), i.o.\right) = 0$, which completes the proof. ∎

For general vector linear systems without full column-ranked C, it is challenging to give necessary conditions for characterizing whether $\limsup_{k \to \infty} \mathrm{Tr}(P_k)$ is bounded or not. In what follows, we will provide with a sufficient condition for the so-called non-degenerate systems, the definition of which is originated from [13].

Definition 1: Consider a system (A, C) in diagonal standard form, i.e., $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $C = [C_1, \ldots, C_n]$. An quasi-equiblock of the system defined as a subsystem $(A_{\mathcal{I}}, C_{\mathcal{I}})$, 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 2: A diagonalizable system (A, C) 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.

Theorem 4: Consider system (1) and (2), and an independent packet loss process $\{\gamma_k\}_{k\in\mathbb{N}_+}$ with a sequence of means $\{\lambda_k\}_{k\in\mathbb{N}_+}$ monotonic. If the system is non-degenerate and there exists an $\mathbb{I}\in\mathbb{N}_+$ such that $\sum_{k=1}^{\infty}(1-\lambda_k)^{\mathbb{I}}<\infty$, then there exists a constant $M(\mathbb{I})$ such that $\mathbb{P}\left(\limsup_{k\to\infty}\mathrm{Tr}(P_k)<M(\mathbb{I})\right)=1$ holds for any $\Sigma_0\geq 0$.

Proof: We first introduce a sequence of stopping time $\{t_j\}_{j\in\mathbb{N}_+}$ as a sequence of consecutive packet arrival times in the following:

$$
\begin{aligned}
t_0 &\triangleq 0,\\
t_1 &\triangleq \min\{k:k\geq 1,\gamma_k=1\},\\
&\vdots\\
t_j &\triangleq \min\{k:k>t_{j-1},\gamma_k=1\}.
\end{aligned}
$$

If $\max\{j:t_j\leq k \text{ and } t_{j+1}>k\}\geq n$, it means that the estimator has received more than n packets from the sensor up to time k. If so, we define that

$$
\begin{aligned}
\tau_1 &\triangleq k-t_i \text{ where } i=\max\{j:t_j\leq k \text{ and } t_{j+1}>k\},\\
\tau_j &\triangleq t_{i-j+2}-t_{i-j+1} \text{ for } 2\leq j\leq n,\\
\tau_j &\triangleq \infty \text{ for } j\geq n+1.
\end{aligned}
$$

To get the desired result, we need the following lemma

Lemma 6: Consider a linear system described by (1) and (2). If $\max\{j:t_j\leq k \text{ and } t_{j+1}>k\}\geq n$ and the system is non-degenerate, then the following inequality holds:

$$
P_k\leq\alpha\prod_{j=1}^{n}(|\lambda_1|+\epsilon)^{2\tau_j},
$$

where α is a constant independent of τ_j and ϵ can be arbitrarily small.

Proof: It is straightforward from Theorem 4 in [13] and the fact that $|\lambda_1|\geq\cdots\geq|\lambda_n|$. ∎

If there exists an $\mathbb{I}\in\mathbb{N}_+$ such that $\sum_{k=0}^{\infty}(1-\lambda_k)^{\mathbb{I}}<\infty$, then we can find a sufficient large positive number $M(\mathbb{I})$ satisfying

$$
M(\mathbb{I})>\alpha|\lambda_1|^{n+\mathbb{I}-2},
$$

which gives,

$$
M(\mathbb{I})>\alpha\prod_{j=1}^{n}(|\lambda_1|+\epsilon)^{2\tau_j}
$$

for a small ϵ and any $\sum_{j=1}^{n}\tau_j\leq n+\mathbb{I}-2$. Given a time index $k\geq n+\mathbb{I}-2$, we can compute that

$$
\begin{aligned}
&\mathbb{P}\left(\mathrm{Tr}(P_k)>M(\mathbb{I})\right)\\
\leq\ &\mathbb{P}\left(\mathrm{Tr}(P_k)>\alpha\prod_{j=1}^{n}(|\lambda_1|+\epsilon)^{2\tau_j}\right),\forall\sum_{j=1}^{n}\tau_j\leq n+\mathbb{I}-2\\
\leq\ &\mathbb{P}\Big(\text{less than } n \text{ packets received between}\\
&\qquad\qquad\qquad \text{time } k-n-\mathbb{I}+2 \text{ and } k\Big)\\
\leq\ &\sum_{j=0}^{n-1}\binom{n+\mathbb{I}-1}{j}\max\{\lambda_{k-n-\mathbb{I}+2},\lambda_k\}^j\\
&(1-\min\{\lambda_{k-n-\mathbb{I}+2},\lambda_k\})^{n+\mathbb{I}-j-1}\\
\leq\ &\sum_{j=0}^{n-1}\binom{n+\mathbb{I}-1}{j}(1-\min\{\lambda_{k-n-\mathbb{I}+2},\lambda_k\})^{\mathbb{I}},
\end{aligned}
$$

where the second last inequality is from the monotonicity of $\{\lambda_k\}_{k\in\mathbb{N}_+}$. Thus,

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

By Lemma 3, its holds that $\mathbb{P}\left(\mathrm{Tr}(P_k)>M(\mathbb{I}),i.o.\right)=0$ even the set of events $\{\mathrm{Tr}(P_k)>M(\mathbb{I})\}_{k\in\mathbb{N}_+}$ are not independent. The proof is complete. ∎

Remark 2: If there exists an $\mathbb{I}\in\mathbb{N}_+$ such that $\sum_{k=1}^{\infty}(1-\lambda_k)^{\mathbb{I}}<\infty$ as assumed in Theorem 4, then we have $\mathrm{Tr}\left(\mathbb{E}[P_k]\right)$. We can argue as follows. A necessary condition for $\sum_{k=1}^{\infty}(1-\lambda_k)^{\mathbb{I}}<\infty$ is $\lim_{k\to\infty}\lambda_k=1$. Consequntly, there exists a sufficient large integer N such that, for any $k\geq N$, $\lambda_k>\lambda_c$, where λ_c is a critical value such that $\mathbb{E}[P_k]\leq M_{\Sigma_0}$ for $\lambda_c<\lambda\leq 1$ and $\forall\Sigma_0\geq 0$ [5].

IV. CONCLUSION

In this paper, we investigate the probabilistic convergence of Kalman filtering with nonstationary packet losses that are modeled by a sequence of independent, but not necessarily identical Bernoulli random variables. We provide some results on how the convergence of the prediction error covariance matrix depends on the statistical property of the nonstationary packet dropout process. Future work includes the expectation bounds, weak convergence and ergodicity properties of discrete-time Kalman filtering with nonstationary random packet losses.

APPENDIX

Proof of Lemma 5: According to the fact $\overline{P}\leq\overline{M}$ that is derived from the definition of $\overline{M}$, one can directly verify that $\overline{\overline{\mathbb{I}}}(M)\leq\underline{\mathbb{I}}(M)$. To show that $\overline{\overline{\mathbb{I}}}(M)$ and $\underline{\mathbb{I}}(M)$ are both finite for any $M\geq\overline{M}$, it suffices to show that there exists

an integer $k \in \mathbb{N}_+$ satisfying $\mathrm{Tr}\left(h^k(\overline{P})\right) > M$. Before proceeding, we introduce the following lemma

Lemma 7: Suppose the system is given by (1) and (2). Then, for any matrix $X \in \mathbb{S}_+^n$, it holds that $\mathrm{Tr}\left(h^k(X)\right) \geq \alpha|\lambda_1(A)|^{2k}$, where $\alpha > 0$ is a constant.

Proof: According to the controllability of $(A, \sqrt{Q})$ we assume, one has that $V \triangleq h^n(0) > 0$. Suppose that there exists a real number $\alpha_0 > 0$ so that $V \geq \alpha_0 I$. Then, for any $k > n$, $h^k(0) \geq \alpha_0 A^{k-n}(A')^{k-n}$ holds. Let us denote the Schur's unitary triangularization of A as $A = UTU^*$ where U is a unitary matrix 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 symmetric and positive semi-definite, one obtains that $\lambda_1\left(A^{k-n}(A')^{k-n}\right)$ is real and that

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

Therefore, we obtain that $\mathrm{Tr}\left(h^k(0)\right) \geq \alpha_n|\lambda_1(A)|^{2k}$ for any $k \geq n$ where $\alpha_n \triangleq \alpha_0|\lambda_1(A)|^{-2n}$. As for $k = 1, \ldots, n-1$, we can take a sequence of positive real numbers, denoted by $\{\alpha_k\}_{1 \leq k \leq n-1}$, such that $\mathrm{Tr}\left(h^k(0)\right) \geq \alpha_k|\lambda_1(A)|^{2k}$ holds. Taking $\alpha = \min\{\alpha_k : k = 1, \ldots, n\} > 0$, we can conclude that $\mathrm{Tr}\left(h^k(X)\right) \geq \mathrm{Tr}\left(h^k(0)\right) \geq \alpha|\lambda_1(A)|^{2k}$, which is the desired result. ∎

From Lemma 7, there always exists an $\alpha > 0$ such that $\mathrm{Tr}\left(h^k(X)\right) \geq \alpha|\lambda_1(A)|^{2k}$. Therefore, as long as we take $k \geq \lceil \frac{\log M - \log \alpha}{2 \log |\lambda_1(A)|} \rceil + 1$, the equality $\alpha|\lambda_1(A)|^{2k} > M$ holds, and the desired result follows. ∎

REFERENCES

[1] T. Hsieh, "Using sensor networks for highway and traffic applications," *IEEE Potentials*, vol. 23, no. 2, pp. 13–16, 2004.
[2] J. Braun, "Intelligent building systems - past, present, and future," in *Proceedings of the 2007 American Control Conference*, New York City, USA, July 2007, pp. 4374–4381.
[3] A. Casavola and G. Franze, "Coordination strategies for networked control systems: A power system application," in *Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision*, Hanoi, Vietnam, December 2008, pp. 503–508.
[4] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, Jan. 2007.
[5] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poola, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, 2004.
[6] X. Liu and A. Goldsmith, "Kalman filtering with partial observation losses," in *Proceedings of the 43rd IEEE Conference on Decision and Contr.*, vol. 4, Dec 2004, pp. 4180– 4186.
[7] Y. Mo and B. Sinopoli, "Towards finding the critical value for Kalman filtering with intermittent observations," *arXiv preprint arXiv:1005.2442*, 2010.
[8] K. Plarre and F. Bullo, "On Kalman filtering for detectable systems with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 386–390, Feb 2009.
[9] M. Huang and S. Dey, "Stability of Kalman filtering with Markovian packet losses," *Automatica*, vol. 43, pp. 598–607, 2007.
[10] L. Xie and L. Xie, "Stability of a random Riccati equation with Markovian binary switching," *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1759–1764, 2008.
[11] 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, 2011.
[12] L. Shi, M. Epstein, and R. M. Murray, "Kalman filtering over a packet-dropping network: A probabilistic perspective," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 594–604, 2010.
[13] Y. Mo and B. Sinopoli, "Kalman filtering with intermittent observations: Tail distribution and critical value," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 677–689, March 2012.
[14] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Transactions on Communications*, vol. 39, no. 5, pp. 653–663, May 1991.
[15] Q. Zhang and S. Kassam, "Hybrid ARQ with selective combining for fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 5, pp. 867–880, May 1999.
[16] M. Stojanovic, "Underwater acoustic communications: Design considerations on the physical layer," in *Proceedings of the 5th Annual Conference on Wireless on Demand Network Systems and Services*, Jan 2008, pp. 1–10.
[17] B. Anderson and J. Moore, *Optimal Filtering*. Prentice Hall, 1979.
[18] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2010, vol. 3.
[19] W. Rudin, *Real and complex analysis*. Tata McGraw-Hill Education, 1987.