European Journal of Control 50 (2019) 62–71

Contents lists available at ScienceDirect

European Journal of Control

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

Heavy-tails in Kalman filtering with packet losses

Matthias Pezzutto[a], Luca Schenato[a], Subhrakanti Dey[b,*]

[a] Department of Information Engineering, University of Padova, via Gradenigo 6/b, Padova 35131, Italy
[b] Institute for Telecommunications Research, University of South Australia, South Australia, Australia

ARTICLE INFO

Article history:
Received 3 October 2018
Revised 23 January 2019
Accepted 10 May 2019
Available online 16 May 2019

Recommended by Prof. T. Parisini

Keywords:
Kalman filtering
Packet losses
Heavy-tailed distributions
Power-law tails

ABSTRACT

In this paper, we study the existence of a steady-state distribution and its tail behaviour for the estimation error arising from Kalman filtering for unstable linear dynamical systems. Although a large body of literature has studied the problem of Kalman filtering with packet losses in terms of analysis of the second moment, no study has addressed the actual distribution of the estimation error. First we show that if the system is strictly unstable and packet loss probability is strictly less than unity, then the steady-state distribution (if it exists) must be heavy tail, i.e. its absolute moments beyond a certain order do not exist. Then, by drawing results from Renewal Theory, we further provide sufficient conditions for the existence of such stationary distribution. Moreover, we show that under additional technical assumptions and in the scalar scenario, the steady-state distribution of the Kalman prediction error has an asymptotic power-law tail, i.e. $P[|e| > s] \sim s^{-\alpha}$, as $s \to \infty$, where α can be explicitly computed. We further explore how to optimally select the sampling period assuming an exponential decay of packet loss probability with respect to the sampling period. In order to minimize the expected value of the second moment or the confidence bounds, we illustrate that in general a larger sampling period will need to be chosen in the latter case as a result of the heavy tail behaviour.

1. Introduction

The proliferation of wireless communication in the past decades is now been penetrating also the industrial automation sectors, thus drawing the interest of a large body of research which has started to analyze the impact of unreliable communication in control and estimation performance [10,33]. In particular, much attention has been placed on the problem of optimal estimation in the presence of packet losses via Kalman Filtering since [30]. However stability and performance have always been evaluated in terms of the error covariance conditioned on the packet loss sequence, i.e. in terms of the second moment of the estimation error. The first major result was to show the existence of a critical packet loss probability for the boundedness of the expected second moment (averaged over the packet loss process) under an i.i.d. packet loss scenario [30]. Later this analysis has been extended to Markov packet losses [12,31,32], to the computations of upper and lower bounds for such critical packet loss probability [24,28,31], to the existence and type of distribution for the error covariance [4,13]. Specific efforts have been directed to the analysis of the multivariable scenario by determining connections between the critical loss probability for mean square stability and algebraic conditions in terms of detectability [21], eigenvalue cycles [20] and non-degeneracy [18].

In this work we concentrate on a continuous-time stochastic *strictly unstable linear system*, i.e. systems which have at least one positive eigenvalue, which is sampled with a sampling period T. Although most of the recent results have rightfully concentrated on the analysis of the error covariance matrix and functions of its trace, we believe that not enough attention has been directed towards understanding the actual distribution of the prediction/estimation error. In fact, if the error distribution is not Gaussian, the steady-state distribution may exist even if the second moment is unbounded. Indeed, this is the case for the estimation error in Kalman filtering with packet losses as shown later.

This work aims to extend our previous work [5]. The major contribution is twofold. The first is to show that under mild conditions, if the original system is strictly unstable, then any steady-state distribution (if it exists) must be heavy-tail. This result implies that large error deviations are more likely to appear than in the standard Gaussian distribution or in thin-tail distributions in general. The second is that a steady-state distribution exists under much milder conditions on the packet loss probability than second moment stability. Indeed for the special case of a scalar unstable system, if the filtering gain is optimally chosen, a

* Corresponding author. Present address: Hamilton Institute, Maynooth University, Co. Kildare, Ireland.

E-mail addresses: matthias.pezzutto@studenti.unipd.it (M. Pezzutto), schenato@dei.unipd.it (L. Schenato), Subhra.Dey@signal.uu.se (S. Dey).

https://doi.org/10.1016/j.ejcon.2019.05.003
0947-3580/

distribution exists as soon as the packet loss probability is strictly less than unity, which is not the case for second moment stability. In other words, there are scenarios in which the second moment does not exist, but the probability that the estimation error is outside a specified interval decreases to zero as the interval size increases, i.e. the error is bounded in probability. Moreover, by exploiting results developed by the stochastic systems community in the area of Renewal Theory and Random Difference Equations [2,3,6,9,14] since the 70's, it is possible to explicitly characterize the tail distribution of the estimation error of the Kalman filter in the scalar case. More specifically, such a distribution, under some technical conditions such as a *non-arithmetic* support of a parameter involving the Random Difference Equation, can be shown to possess a power-law tail with an explicit characterization of the power exponent α as well as its coefficient c, i.e.

$$\lim_{s\to\infty} \frac{P[|e| > s]}{s^{-\alpha}} = c$$

Similar results were also observed in the context of limited-rate control systems [25], which however uses a somewhat different framework than Kalman Filtering with packet losses. As a simple corollary of these results is that also the distribution of the second moment, i.e. $P[e^2 < E]$, must be heavy-tailed itself under the same conditions mentioned above for the power-law tail for the estimation error distribution, i.e. we recover the same results observed previously in [18]. This also implies that the moments of error covariance are unbounded beyond a certain order, i.e. $\exists m_c > 0$ such that $\mathbb{E}[(e^2)^m] = \infty$ for $m > m_c$, as previously observed in [12]. Another corollary of our work is that, since the error distribution is heavy-tailed even when the system is second-moment stable, the confidence bounds can be rather different from what one would obtain by (incorrectly) assuming the steady-state distribution is Gaussian with variance obtained from the modified Riccati Equation which arises in the context of Kalman filtering with packet losses [30]. In other words, the $3\sqrt{\text{trace}(P)}$ estimation of the confidence bounds, where P is the second order moment of the steady-state error distribution, can be more optimistic than what occurs in reality, i.e. large error values are not rare, and can lead to a very negative impact in safety-critical applications. The later part of the paper explores the optimal sampling of two stochastic continuous-time unstable systems both in terms of minimizing the expected second moment $\bar{p}$ and in terms of the confidence bound for the steady-state error with an approximately 99% confidence probability. We observe, that the optimal sampling period in the latter case is larger that the one dictated by the former. This is indeed another consequence of the heavy-tailed behaviour of the distribution, since it implies that it is better to incur in a larger delay using a larger sampling period than having a more heavy-tailed distribution from a confidence bound perspective.

2. Modeling and definitions

We start with a continuous-time multivariable state space system given by the following stochastic differential equation:

$$dx(t) = Ax(t)dt + Bu(t)dt + d\bar{w}(t) \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the control and $\bar{w}(t)$ is the process noise described by a Wiener process with independent Gaussian distributed increments such that $\bar{w}(t+t') - \bar{w}(t) \sim \mathcal{N}(0, t'Q)$, and $x_0 \sim \mathcal{N}(0, P_0)$. This process is sampled uniformly with a sampling period of T and zero-order hold to produce the following discrete-time system:

$$x_{k+1} = \bar{A}(T)x_k + \bar{B}(T)u_k + w_k \tag{2}$$

where $\bar{A}(T) = e^{AT}$, $\bar{B}(T) = \int_0^T e^{A\tau}Bd\tau$ and w_k is independent and identically Gaussian distributed with zero mean and variance $\overline{Q}(T) = \int_0^T e^{A\tau}Qe^{A^T\tau}d\tau$, i.e. $w_k \sim \mathcal{N}(0, \overline{Q}(T))$. We denote the multivariable discrete-time measurements of this system as $y \in \mathbb{R}^p$, and given by a linear equation:

$$y_k = Cx_k + v_k \tag{3}$$

where $v_k \in \mathbb{R}^p$ is the measurement noise, independent and identically Gaussian distributed with zero mean and variance R, i.e. $v_k \sim \mathcal{N}(0, R)$, independent of w_k, and x_0. These measurements are transmitted to a remote estimator over a lossy channel such that each measurement is either received or lost according to a Bernoulli process $\gamma_k \in \{0, 1\}$, independent of v_k, w_k, x_0, such that the $\mathbb{P}[\gamma_k = 0] = \bar{\gamma}(T)$ is the packet loss probability, hence $\mathbb{E}[\gamma_k] = 1 - \bar{\gamma}(T)$. Naturally, the packet loss probability is a function of the sampling period T, since a high sampling rate results in a higher packet transmission rate and a higher packet loss probability, assuming that all other channel conditions remain unchanged. We assume that $\bar{\gamma}(T)$ is a continuous decreasing function of T. Specific forms of $\bar{\gamma}(T)$ depend on the underlying modulation and coding schemes of the associated communication system.

Based on the information set $\mathcal{Z}_k \triangleq \{y_0, y_1, \dots y_k, \gamma_0, \gamma_1, \dots, \gamma_k\}$, the remote Kalman predictor $\hat{x}_{k+1|k} := \mathbb{E}[x_{k+1} \mid \mathcal{Z}_k]$ is:

$$\hat{x}_{k+1|k} = \bar{A}(T)\hat{x}_{k|k-1} + \bar{B}(T)u_k + K_k\gamma_k(y_k - C\hat{x}_{k|k-1}) \tag{4}$$

where K_k is the time-varying Kalman prediction gain given by:

$$K_k = AP_kC^T(CP_kC^T + R)^{-1}$$

$$P_{k+1} = AP_kA^T + \overline{Q}(T) - \gamma_k AP_kC^T(CP_kC^T + R)^{-1}CP_kA^T$$

and $P_0 = \mathbb{E}[x_0x_0^T]$ as shown in [30]. We then define the prediction error $e_k := x_k - \hat{x}_{k|k-1}$ whose dynamics is given by:

$$e_{k+1} = (\bar{A} - \gamma_k K_k C)e_k + (w_k - \gamma_k K_k v_k). \tag{5}$$

In the following dissertation, some results are obtained and outlined for the special case of the scalar systems. To highlight this particular case, the involved quantities will be denoted in lowercase: the system will be indicated by a, b and by $\bar{a}$, $\bar{b}$, c, for the continuous-time and for the discrete-time case, respectively; the variance of the Gaussian increments of the Wiener process will be q, while the variance of the process noise and of the measurement noise of the discrete-time system will be σ^2 and σ_v^2, respectively; finally, the prediction gain will be denoted by $\bar{k}$ and the error variance by $\bar{p}(T)$.

In the next sections we will study the probability distribution of the random vector e_k and in particular its steady-state distribution. To characterize such a distribution we first need to introduce a few definitions.

To this end, we denote the set of integers by $\mathbb{Z}$, and the set of real numbers by $\mathbb{R}$. For a random variable X, X_+, X_- denote the positive and negative part of the random variable, respectively. Moreover $\log^+ x = \log x$ for $x \geq 1$ and 0 for $x \in (0, 1)$. In what follows, $\|\cdot\|$ indicates the 2-norm for vectors or the induced 2-norm for matrices.

Definition 1. Let us consider the multivariate random variable $X \in \mathbb{R}^\ell$. We say that its distribution is heavy-tailed if there exists a $r > 0$ such that $\mathbb{E}[\|X\|^r] = +\infty$.

Definition 2. Let us consider the multivariate random variable $X \in \mathbb{R}^\ell$. We say that its distribution is power-law if there exists $\alpha > 0$, $c > 0$ such that $\lim_{L\to\infty} P[\|X\| > L] \sim cL^{-\alpha}$.

We will also exploit the following moment inequality (see [15] p. 263 for a proof):

Theorem 1. *Let us consider the multivariate random variables $X, Y \in \mathbb{R}^\ell$ If $E[\|X\|^r] < \infty$, $E[\|Y\|^r] < \infty$ and $E[Y|X] = 0$ almost sure, then:*

$$E[\|X + Y\|^r] \geq E[\|X\|^r], \ \forall r \geq 1. \tag{6}$$

3. Heavy tail properties of prediction error

We now show that if the original system is strictly unstable and if the process noise excites the unstable modes, then the steady-state distribution, if it exists, must be heavy-tail. Sufficient conditions for the existence of such a steady-state distribution are addressed in the next section since they require additional conditions on the error dynamics.

Theorem 2. *Consider the error dynamics given by (5). Assume that* $\mathbb{E}[\|x_0\|^r] < \infty$, $\mathbb{E}[\|w_k\|^r] < \infty$, $\mathbb{E}[\|v_k\|^r] < \infty, \forall r \geq 1$, *and that the pair* $(\bar{A}(T), \overline{Q}(T))$ *is reachable. If* $\bar{A}(T)$ *is strictly unstable,* $0 < \bar{\gamma}(T) < 1$, *and there exists a steady-state distribution for* e_k, *then it must be heavy-tail.*

Proof. For readability purposes, we drop the dependence in T of the various parameters, i.e. $\bar{A}(T) = \bar{A}$, $\bar{\gamma}(T) = \bar{\gamma}$, $\overline{Q}(T) = \overline{Q}$. Consider the error dynamics given by (5) and the following dynamics without noise:

$$\widetilde{e}_{k+1} = (\bar{A} - \gamma_k K_k C)\widetilde{e}_k, \ \ \widetilde{e}_1 = e_1, \ \ \forall k > 1.$$

Let $X = (\bar{A} - \gamma_k K_k C)e_k$ and $Y = (w_k - \gamma_k K_k v_k)$. Now

$$\mathbb{E}[Y \mid X] = \overline{\gamma}\mathbb{E}[w_k \mid \bar{A}e_k] + (1 - \overline{\gamma})\mathbb{E}[w_k - K_k v_k \mid (\bar{A} - K_k C)e_k] = 0$$

since w_k, v_k are zero mean and independent of $\gamma_{k-1}, \dots \gamma_0$ and e_k. We can use this observation to apply Lemma 1 inductively to get:

$$\mathbb{E}[\|e_k\|^r] \geq \mathbb{E}[\|\widetilde{e}_k\|^r], \ \ \forall r \geq 1, \ \ \forall k \geq 1$$

therefore we can restrict our analysis to $\widetilde{e}_k$. We now compute the expectation of $\widetilde{e}_k$ with respect to γ_k assuming all the other random variables fixed:

$$\begin{aligned}\mathbb{E}[\|\widetilde{e}_{k+1}\|^r] &= \mathbb{E}[(\widetilde{e}_{k+1}^T\widetilde{e}_{k+1})^{\frac{r}{2}}] = \bar{\gamma}\mathbb{E}[(\widetilde{e}_k^T\bar{A}^T\bar{A}\widetilde{e}_k)^{\frac{r}{2}}] + \\ &\quad + (1-\bar{\gamma})\mathbb{E}[(\widetilde{e}_k^T(\bar{A}-K_kC)^T(\bar{A}-K_kC)\widetilde{e}_k)^{\frac{r}{2}}] \\ &\geq \bar{\gamma}\mathbb{E}[(\widetilde{e}_k^T\bar{A}^T\bar{A}\widetilde{e}_k)^{\frac{r}{2}}].\end{aligned}$$

By induction, using the same argument, we have:

$$\begin{aligned}\mathbb{E}[\|\widetilde{e}_{k+1}\|^r] &\geq \bar{\gamma}^k\mathbb{E}[(\widetilde{e}_1^T(\bar{A}^T)^k\bar{A}^k\widetilde{e}_1)^{\frac{r}{2}}] \\ &= \bar{\gamma}^k\mathbb{E}[(\mathrm{trace}((\bar{A}^T)^k\bar{A}^k\widetilde{e}_1\widetilde{e}_1^T))^{\frac{r}{2}}] \\ &\geq \bar{\gamma}^k(\mathrm{trace}(\bar{A}^k\mathbb{E}[\widetilde{e}_1\widetilde{e}_1^T](\bar{A}^T)^k))^{\frac{r}{2}}, \quad r \geq 2\end{aligned}$$

where the last step has been obtained using Jensen inequality which holds for $r \geq 2$. Now, by construction $\mathbb{E}[\widetilde{e}_1\widetilde{e}_1^T] = \mathbb{E}[e_1 e_1^T]$. Using this notation we have:

$$e_1 = (\bar{A} - \gamma_0 K_0 C)e_0 + w_0 - \gamma_0 K_0 v_0.$$

If we define $X = w_0$ and $Y = (\bar{A} - \gamma_0 K_0 C)e_0 - \gamma_0 K_0 v_0$, these random variables satisfy Lemma 1, therefore we have

$$\mathbb{E}[e_1 e_1^T] \geq \mathbb{E}[w_0 w_0^T] = Q$$

and consequently

$$\mathbb{E}[\|\widetilde{e}_{k+1}\|^r] \geq \bar{\gamma}^k(\mathrm{trace}(\bar{A}^kQ\bar{A}^k))^{\frac{r}{2}}.$$

Now note that $\bar{A}^kQ\bar{A}^k = \mathcal{L}^k(Q)$, where $\mathcal{L}(\bar{A}X\bar{X}^T\bar{A}^T)$ is the Lyapunov operator whose eigenvalues are $\lambda_i\lambda_j^*$ where λ_i, λ_j are the eigenvalues of $\bar{A}$. Since the pair $(\bar{A}, \overline{Q})$ is reachable, then Q surely excites the largest eigenvalue $|\lambda_{max}|^2$, where $|\lambda_{max}| > 1$. Therefore there exists a $\bar{k} \in \mathbb{N}$ and $c > 0$ such that:

$$\mathrm{trace}(\bar{A}^kQ\bar{A}^k) \geq c|\lambda_{max}|^{2k}, \quad k \geq \bar{k}.$$

This in turns implies that:

$$\mathbb{E}[\|\widetilde{e}_{k+1}\|^r] \geq c^{\frac{r}{2}}\bar{\gamma}^k|\lambda_{max}|^{rk} = c^{\frac{r}{2}}(\bar{\gamma}|\lambda_{max}|^r)^k.$$

Now if we choose $\bar{r} := \max\{2, -\frac{\log\bar{\gamma}}{\log|\lambda_{max}|}\} < \infty$ for $\bar{\gamma} \neq 1$, then for each $r > \bar{r}$

$$\lim_{k\to\infty}\mathbb{E}[\|\widetilde{e}_k\|^r] = +\infty.$$

□

Remark 1. Note that the previous proof does not require K_k to be the optimal Kalman gains, but holds for any choice of such gains, i.e. there is no linear estimation strategy that can avoid the heavy-tail property of the distribution if the system is strictly unstable. If the system is strictly stable, i.e. $|\lambda_{max}| < 1$, then it is possible to show that if a steady-state distribution exists, it must be thin-tail. If the system is only marginally stable, i.e. $|\lambda_{max}| = 1$, then determining the behaviour of the tail distribution is much more challenging even for scalar systems (see for example [11]) and it is outside the scope of this work.

4. Existence of a steady-state distribution in constant-gain filtering

The optimal Kalman filter requires time-varying gains which depends on the packet loss sequence [30], therefore determining even the existence of a steady-state distribution is challenging in this scenario. Differently, we will make use of constant gain filter which will make the subsequent analysis simpler. More specifically we will consider the following error dynamics:

$$e_{k+1} = (\bar{A} - \gamma_k\bar{K}C)e_k + (w_k - \gamma_k\bar{K}v_k) \tag{7}$$

where $\bar{K}$ does not depend on time k and to simplify the notation we drop the dependency on T of $\bar{A}$. A possible choice of $\bar{K}$ is discussed in [27] for example, where it was shown that under the condition that $\lim_{k\to\infty}\mathbb{E}(e_ke_k^T) < \infty$, $\bar{K}$ can be chosen to be the gain that minimizes $\lim_{k\to\infty}E(e_ke_k^T)$, and this choice results in a marginal performance loss in terms of expected estimation error variance. Indeed,

$$\bar{K} = \bar{A}PC^T(CPC^T + R)^{-1}$$

where $P \triangleq \lim_{k\to\infty}\mathbb{E}(e_ke_k^T)$ satisfies the modified algebraic Riccati equation:

$$P = \bar{A}P\bar{A}^T + \overline{Q} - (1-\bar{\gamma})\bar{A}PC^T(CPC^T + R)^{-1}CP\bar{A}^T. \tag{8}$$

We will now provide sufficient conditions for the existence of the steady-state distribution. The following theorems are mostly based on the theory of linear difference equations with stochastic coefficients [14] and on the theory of Lyapunov exponent for products of random matrices [8] which are summarized in the following Lemma:

Lemma 3. *Let us consider the following stochastic dynamic system*

$$E_k = F_kE_{k-1} + Z_k$$

where $E_k, Z_k \in \mathbb{R}^n$, $F_k \in \mathbb{R}^{n\times n}$ *are random variables and the pairs* $(E_k, Z_k), k \geq 1$, *are independent and identically distributed (i.i.d). Let us also define the following random variables*

$$J_k := \sum_{n=1}^{k} F_1 \dots F_{n-1}Z_n, \quad J := \sum_{n=1}^{\infty} F_1 \dots F_{n-1}Z_n$$

and the so called Lyapunov exponent:

$$\bar{\lambda} = \lim_{k\to+\infty}\frac{1}{k}\log(||F_1 \dots F_k||)$$

Then the following properties hold true:

(i) *The random variable* E_k *has the same distribution as* $J_k + F_kF_{k-1}\dots F_1E_0$, *for a given* E_0.
(ii) *If* $\mathbb{E}(\log^+||F_1||) < \infty$, *then* $\bar{\lambda}$ *exists with probability 1 (w.p.1) (not necessarily finite) and it is equal to*

$$\bar{\lambda} = \lim_{k\to+\infty}\frac{1}{k}\mathbb{E}[\log(||F_1 \dots F_k||)] \leq \mu := \mathbb{E}[\log\|F_1\|]$$

(iii) *If there exists* $r > 0$ *such that* $\mathbb{E}(\|Z_1\|^r) < \infty$, *and if* $\bar{\lambda} < 0$ *w.p.1, then the distribution of* J_k *converges w.p.1 to the distribution J.*
(iv) *If* $\bar{\lambda} < 0$ *w.p.1, then* $F_n F_{n-1} \dots F_1 E_0$ *converges to zero exponentially.*

The previous results can be combined to obtain the following lemma:

Lemma 4. *Assume that* $(F_k, Z_k), k \geq 1$, *are i.i.d and* $\exists r > 0$ *such that* $\mathbb{E}(\|Z_1\|^r) < \infty$. *If* $\bar{\lambda} < 0$ *w.p.1, then the distribution of* E_k *converges w.p.1 to the distribution J independently on* E_0.

We are now ready to prove the main theorem in this section:

Theorem 5. *Let us consider the stochastic dynamical system (7). Let* $w_k \sim \mathcal{N}(0, \bar{Q})$, $v_k \sim \mathcal{N}(0, R)$, $x_0 \sim \mathcal{N}(0, P_0)$ *to be mutually independent, and i.i.d. white. Let* γ_k *be an i.i.d. Bernoulli random variable with* $\bar{\gamma} := \mathbb{P}[\gamma_k = 0]$. *If*

$$\mu := \bar{\gamma}\log(||\bar{A}||) + (1-\bar{\gamma})\log(||\bar{A} - \bar{K}C||) < 0 \qquad (9)$$

then e_k *converges in probability to a steady-state distribution.*

Proof. The proof is obtained by verifying the hypotheses of Lemma 4 where $F_k = (\bar{A} - \gamma_k \bar{K} C)$ and $Z_k = (w_k - \gamma_k \bar{K} v_k)$. Clearly (F_k, Z_k) are i.i.d. random vectors. Under hypothesis (9), the Lyapunov exponent $\bar{\lambda}$ is negative, in fact

$$\bar{\lambda} \leq \mu = \mathbb{E}[\log\|F_1\|] = \overline{\gamma}\log(||\bar{A}||) + (1-\overline{\gamma})\log(||(\bar{A} - KC||)$$

At this point we just need to show that $\exists r > 0$ such that $\mathbb{E}(\|Z_1\|^r) < \infty$. Let us pick $r = 2$, therefore

$$\mathbb{E}\big(\|Z_1\|^2\big) = \mathbb{E}\big[Z_1^T Z_1\big] = \mathbb{E}\big[\text{trace}(Z_1 Z_1^T)\big] = \text{trace}\big(\mathbb{E}\big[Z_1 Z_1^T\big]\big)$$
$$= \text{trace}\big(\mathbb{E}\big[w_k w_k^T + \gamma_k^2 K v_k v_k^T K^T + \gamma_k w_k v_k^T K^T + \gamma_k K v_k w_k^T\big]\big)$$
$$= \text{trace}(\bar{Q} + (1-\overline{\gamma})\bar{K}R\bar{K}^T) < \infty$$

which concludes the proof. □

The previous theorem provides only a sufficient condition for the existence of a steady-state distribution which is basically based on the observation that the scalar $\mu = \mathbb{E}[\log\|F_1\|]$ is an upper bound for the exact Lyapunov exponent $\bar{\lambda}$. Such a condition is not very tight in general, and in the past fifty years a large body of works has addressed the problem of finding exact or tight lower and upper bounds for the Lyapunov exponent for certain classes of product of random matrices [8,23]. It has been shown that even in the case of products of 2×2 random matrices, explicit computation of the Lyapunov exponent is only possible in some very special cases where the random matrices have specific structural properties and distributions [16]. It is also important to observe that the Lyapunov exponent does not depend on the specific norm adopted, therefore the 2-norm adopted in inequality (9) can be substituted with any other norm in order to find tighter bounds on the Lyapunov exponent. As expected, the Lyapunov exponent can be computed explicitly for scalar systems as reported in the following corollary.

Corollary 6. *Let us consider the conditions in Theorem 5. If the system is scalar, then the Lyapunov exponent is given by*

$$\bar{\lambda} = \mu := \bar{\gamma}\log(|\bar{a}|) + (1-\bar{\gamma})\log(|\bar{a} - \bar{k}c|).$$

In the scalar scenario, the previous corollary suggests some further considerations. First, in the case without measurement noise, if $\bar{k}$ is optimally chosen, i.e. $\bar{k} = a/c$ according to (4), we have that $\bar{\lambda} \to -\infty < 0$, i.e. the steady-state distribution exists, as long as $\bar{\gamma} < 1$. Moreover, it allows us also to compare the stability region for which the error dynamics has a steady-state distribution, i.e. the error is bounded in probability, and when the error dynamics is mean square stable, i.e. when its second moment is bounded. In the latter case, the mean square stability condition can be found in [27] and is given by:

$$\overline{\gamma}\bar{a}^2 + (1-\overline{\gamma})(\bar{a} - \bar{k}c)^2 < 1. \qquad (10)$$

Obviously second moment stability is a sufficient condition for stability in probability. If the error distribution were Gaussian, then second moment stability and stability in probability would coincide, however, this is not the case in the problem at hand since, according to the results of the previous section, if such a steady-state distribution exists, it is heavy-tailed and therefore not Gaussian. It follows that even when the packet loss is above the threshold for the second moment stability, the error can be bounded in probability. This is indeed confirmed by the observation that condition (9) is less stringent that the condition (10). Just as an example Fig. 1 illustrates the stability region as a function of packet loss $\bar{\gamma}$ and filter gain $\bar{k}$ for the two aforementioned conditions.

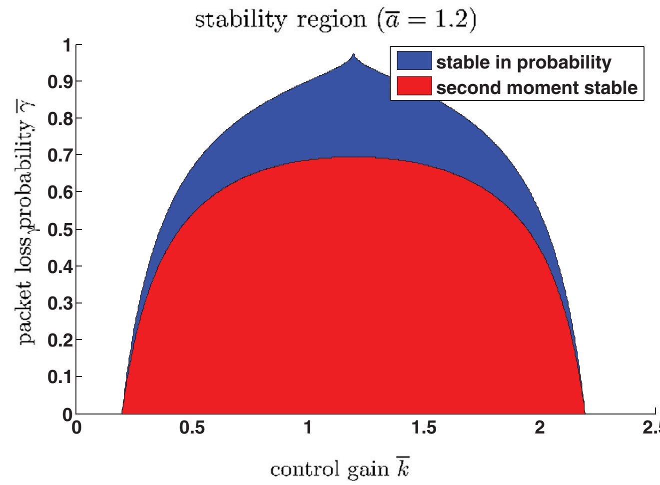

Fig. 1. Plot comparing the stability region in terms of second moment (red) and in probability (blue) as a function of packet loss and filter gain for $c = 1$ and $\bar{a} = 1.2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Computation of the stationary distribution

For simplicity, we still consider the scalar case, and we denote with $\delta^2 = \sigma^2 + \bar{k}^2\sigma_v^2$ the variance of $w_k - \bar{k}v_k$. Note however, that the expression for the steady-state distribution of the prediction error e_k presented below can be easily extended to the multivariable case. Under the assumptions of Theorem 5 such that the steady-state distribution exists, let us denote the stationary distribution of e_k in (7) as $g_\infty(e)$. Then $g_\infty(e)$ satisfies the following integral equation:

$$g_\infty(z) = \bar{\gamma}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(z-\bar{a}e)^2}{2\sigma^2}}g_\infty(e)de$$
$$+(1-\bar{\gamma})\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\delta^2}}e^{-\frac{(z-(\bar{a}-\bar{k})e)^2}{2\delta^2}}g_\infty(e)de. \qquad (11)$$

The above result follows easily from basic probability theory by deriving the cumulative distribution function of the stationary distribution conditioning on the two values of γ_k and then taking the derivative to obtain $g_\infty(e)$. Details are omitted as this derivation is elementary.

While there is no exact closed form solution to the above equation to the best of the authors' knowledge, the easiest way to find an approximate solution to (11) is to discretize the real line for z to denote $g_\infty(z)$ by a finite length vector $\bar{g}$, and replace the integrals on the right hand side by matrix vector products. This then

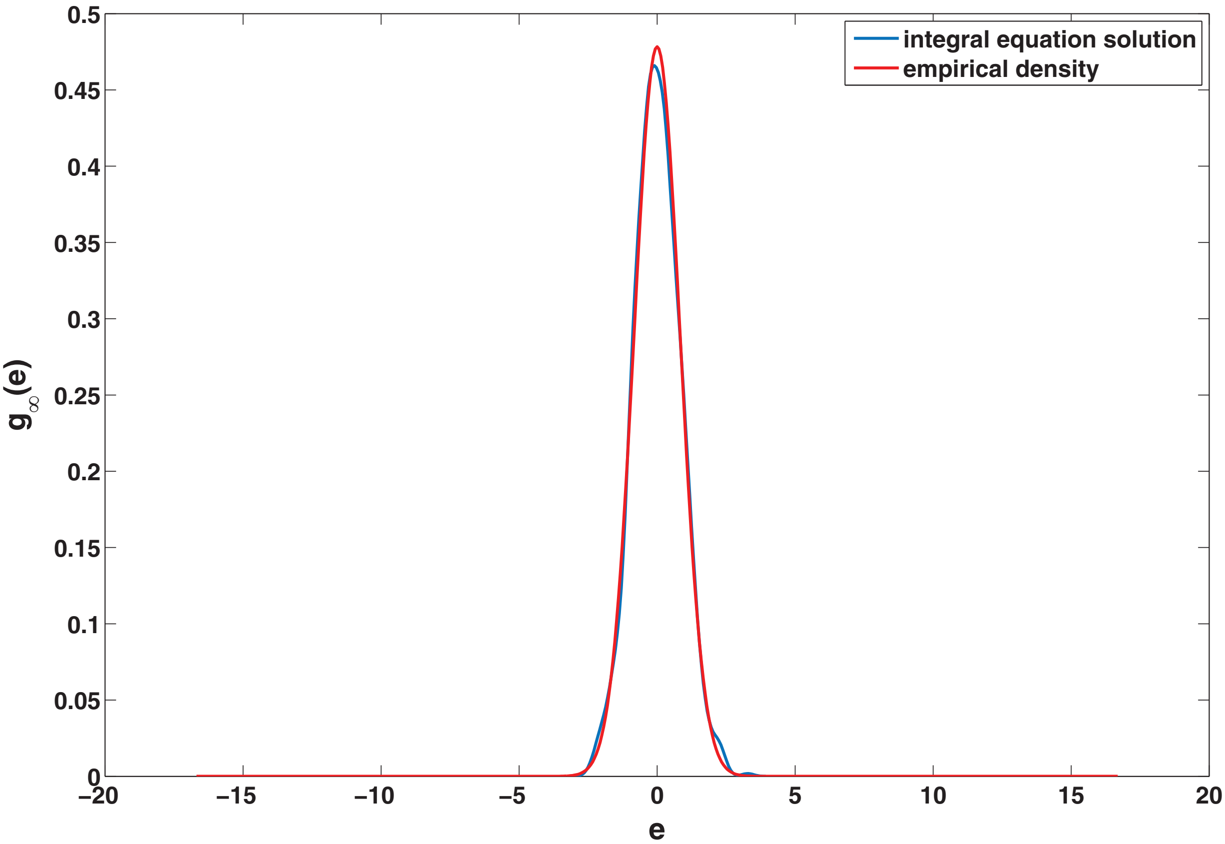

Fig. 2. Plot comparing the stationary distribution of the prediction error obtained via approximate solution to the integral equation (blue) and an empirical method (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

leads to a linear equation involving $\bar{g}$, which can be solved iteratively until the solution converges within a given tolerance. As always, the larger the number of discretization points (the longer the vector $\bar{g}$), the better the approximation. Fig. 2 below compares the stationary distribution of the prediction error for a system with $\bar{a} = 1.9251,\ \bar{k} = 1.4164,\ \sigma_v^2 = 0.25,\ \sigma^2 = 0.6960$, corresponding to a sampling period of $T = 6.55$ ms. The blue graph shows the density obtained using the discretized approximation to the integral equation (11) with 500 discretization points, whereas the red plot shows the corresponding density obtained by using the MATLAB kernel smoothing function "ksdensity" using 1000 points of the prediction error generated by Monte Carlo simulations. As can be seen, the approximation to the integral equation is quite close to the empirical density.

5. Asymptotic power-law tails of the prediction error: scalar case

For the purpose of this section, we consider the scalar case with the constant gain Kalman filter, where $e_k \in \mathbb{R}$, and we rewrite (7) as $e_{k+1} = f_k e_k + z_k$, noting that $f_k = (\bar{a} - \gamma_k \bar{k} c)$, and $z_k = (w_k - \gamma_k \bar{k} v_k)$. Without loss of generality, we take $c = 1$. We note that since γ_k is i.i.d. Bernoulli, and z_k is i.i.d. Gaussian noise, the random process pair (f_k, z_k) is i.i.d. These types of random difference equations have been studied extensively because of their applications in a number of fields such as economics, finance, evolution modelling, and in general for studying random walks in random environments [14].

For the purpose of explaining the existing mathematical theory behind this random equation, we will rewrite it as a general equation in terms of its stationary state (when the stationary distribution exists) as $E \overset{d}{=} FE + Z$, where $\overset{d}{=}$ denotes equality in distribution, similar to [3,6]. We now assume that (F, Z) is a general i.i.d. random process. Depending on the ranges of values the random process F can take, the stationary distribution of E, when it exists, has different asymptotic tail distributions. In particular, for the case where $F \geq 0$ almost sure (a.s.), it has been shown in a variety of works that under mild conditions on the distribution of Z, one can show that as $s \to \infty$, $P(E > s) \sim Cs^{-\alpha}$ for some $\alpha > 0$. In other words, the stationary distribution of E has an asymptotically power-law tail. This result was first shown in [14], with a number variations appearing thereafter. However, it is the paper [6], that is primarily cited as the one which generalized this result, gave a precise description of tail exponent and the constant factors involved, and provided a less complicated proof using Renewal theory (see Lemma 2.2, Theorem 2.3 and Theorem 4.1). In order to keep the presentation simple, we will quote a version of this Theorem that is available in [3], as Theorem 2.4.4., using the notations of our paper.

Theorem 7 [3]**.** *Assume* (i) $F \geq 0$ *a.s. and* (ii) *the law of* $\log F$ *conditioned on* $F > 0$ *is non-arithmetic.*[1] *Assume also that* (iii) *there exists* $\alpha > 0$ *such that* $\mathbb{E}[F^\alpha] = 1$, $\mathbb{E}[|Z|^\alpha] < \infty$, *and* $\mathbb{E}[F^\alpha \log^+ F] < \infty$. *Finally, assume that (iv)* $P(Fs + Z = s) < 1$ *for all* $s \in \mathbb{R}$.

Then the equation $E \overset{d}{=} FE + Z$ *has a solution* E *which is independent of* (F, Z) *and there exist constants* c_+, c_- *such that* $c_+ + c_- > 0$, *and*

[1] F is non-arithmetic if it is not supported in any of the sets $h\mathbb{Z}$, $h \geq 0$ and $\mathbb{Z}$ denotes the set of integers.

$$P(E > s) \sim c_+ s^{-\alpha},\ P(E < -s) \sim c_- s^{-\alpha},\ s \to \infty,$$

$$c_+ = \frac{\mathbb{E}\left[(FE+Z)_+^\alpha - (FE)_+^\alpha\right]}{\alpha m_\alpha}$$

$$c_- = \frac{\mathbb{E}\left[(FE+Z)_-^\alpha - (FE)_-^\alpha\right]}{\alpha m_\alpha} \tag{12}$$

where $0 < m_\alpha = \mathbb{E}[F^\alpha \log(F)] < \infty$.

Before we discuss the implications of this result in our specific case of (7), we note that under the assumptions (i), (ii) and (iii) of Theorem 7 above, it automatically follows that (i) $-\infty \le E[\log F] < 0$ (thus guaranteeing stationarity) and (ii) $0 < m_\alpha < \infty$ - see Lemma 2.2 of [6]. A proof of the above theorem using Renewal theory is provided also in [3].

Now let us apply Theorem 7 to (7) rewritten as $e_{k+1} = f_k e_k + z_k$. It is clear that $f_k \in \{\bar{a}, \bar{a} - \bar{k}\}$ where $\bar{a} > 1$, and therefore $f_k \ge 0$ can be guaranteed as long as $\bar{k} \le \bar{a}$. The non-arithmetic requirement on the distribution of $\log(f_k)$ essentially means that $\frac{\log(\bar{a})}{\log(\bar{a}-\bar{k})}$ cannot be a rational number, which requires also that $\bar{k} < \bar{a}$. In order to guarantee the existence of the stationary solution as indicated in the previous section, we also need $\bar{a} - \bar{k} < 1$. Thus we need $\bar{a} - 1 < \bar{k} < \bar{a}$: note that, if one chooses $\bar{k}$ the Kalman gain according to (4) with $\overline{p}$ satisfying (8), it is satisfied. For any $\bar{\gamma} < 1$, one can show (since $0 < (\bar{a} - \bar{k}) < 1$, and $\bar{a} > 1$) that there exists an $\alpha > 0$ such that the condition $\mathbb{E}[f_k^\alpha] = 1$ is satisfied. It is easy to check that $\mathbb{E}[f_k^\alpha \log^+ f_k] = \bar{a}^\alpha \log(\bar{a})\bar{\gamma} < \infty$. We note also that all absolute moments of z_k are finite. Finally, we have that, since the stationary distribution of e_k is continuous, assumption (iv) of the above theorem is also satisfied. Therefore all the assumptions of the previous theorem apply and we can conclude that the stationary distribution of e_k satisfies the asymptotic power-law tail behaviour as described in Theorem 7. Since the distribution of e_k is symmetric around the origin for all k, so is $g_\infty(e)$, and therefore the two constants c_+, c_- in (12) are equal and positive, leading to the result $P(|e| > s) \sim 2c_+ s^{-\alpha}$, as $s \to \infty$.

The case where the distribution of $\log f_k$ conditioned on $f_k \ne 0$ is *arithmetic*, the analysis of the tail probability is more complex and was carried out in [7,9]. The basic result is that in this case one can prove [3] that there are constants $c_a > c_b > 0$ such that

$$c_a \le \liminf_{s\to\infty} s^\alpha P(|e| > s) \le \limsup_{s\to\infty} s^\alpha P(|e| > s) \le c_b \tag{13}$$

Remark 2. If one considers the no measurement noise case (with the optimal Kalman filtering gain minimizing the expected prediction error variance, $\bar{k} = \bar{a}$), the prediction error recursion follows $e_{k+1} = \bar{a}(1 - \gamma_k)e_k + w_k$. Then it is seen that in this case, $f_k = \bar{a}$ with probability $\bar{\gamma}$ and 0 with probability $(1 - \bar{\gamma})$. This is a special case of the arithmetic distribution scenario, and it is possible that one can obtain the asymptotic power-law tail properties using simpler arguments that the Renewal theory arguments used in [7,9]. Currently, the authors are pursuing a simple proof of this result and only partial progress has been made. While the exponent of the power-law tail is easy to derive using some basic results from [14], it is the calculation of the coefficients c_a, c_b in (13) that has proved to be difficult.

Remark 3. The power-law exponent α in this case is the solution to the equation:

$$\bar{a}^\alpha \bar{\gamma} + (1 - \bar{\gamma})(\bar{a} - \bar{k})^\alpha = 1.$$

The previous expression shows that such an α exists as long as $\bar{\gamma} < 1$, which implies that such a distribution exists even when the estimator is not mean square stable, i.e. $\bar{\gamma} > \frac{1}{\bar{a}^2}$, and that it is heavy-tailed (power-law) even if the estimator is mean square stable, i.e. $\bar{\gamma} < \frac{1}{\bar{a}^2}$. Clearly, the less unstable the system is and the lower the packet loss probability is, the faster the tail goes to zero, but still remains heavy-tailed.

Remark 4. Another interesting observation is that if $\bar{k}$ is to be chosen to maximize α, so that the asymptotic decay rate of the power-law tail is the fastest, then one can easily check that $\bar{k}$ must be chosen as $\bar{a}$, such that:

$$\alpha = \frac{\log\left(\frac{1}{\bar{\gamma}}\right)}{\log \bar{a}}.$$

This illustrates that the gain $\bar{k}$ that maximizes the asymptotic power-law decay rate of the tail of the prediction error is not the same as the one that minimizes the second moment.

Fig. 3 illustrates a plot of the $\log(P(e > s))$ versus $\log(s)$ for the following parameters: $\bar{a} = e^{\sqrt{\frac{1}{17}}}$,[2] $\bar{k} = 0.7$, $\sigma^2 = 0.01$, $\sigma_v^2 = 0.1$, and $\bar{\gamma} = 0.4$. From this plot one can calculate an estimated value of the power-law exponent α as 3.4692, whereas its theoretically obtained value is 3.3792. Similarly the constant c_+ is estimated to be 0.0223, whereas its theoretically obtained value from (12) is given by 0.0270.

Similarly, Fig. 4 illustrates a plot of $\log(P(e > s))$ versus $\log(s)$ in the no measurement noise case, with $\bar{a} = 1.3$, $\sigma^2 = 0.01$ and $\bar{\gamma} = 0.4$. In this case, the theoretically calculated value of the power-law exponent is $\alpha = 3.4924$, whereas the one estimated from this plot is 3.3846.

Remark 5. It should be noted that the asymptotic power-law tail results hold also for random difference equations involving vector processes as well as where the coefficients are Markovian rather than i.i.d. - see for example [26], and Theorem 6 in [14]. However, the technical conditions under which these results hold are significantly more complicated, especially the conditions involving the non-arithmetic distributions. Extension of the asymptotic power-law tail results to the multivariable case is therefore left for future work.

6. Implications of heavy tail properties: choice of sampling rate in confidence bounds versus second moment stability

As discussed earlier, majority of the literature on Kalman filtering with packet loss has focused on analysing the estimation error covariance, its expected value [30] or tail bounds such as in [29]. In [18], a power-law tail behaviour for the solution to the random Riccati equation was established. Different to these works, here we focus on the estimation/prediction error itself and its tail behaviour. We consider the constant gain Kalman filter based error dynamics (7), and investigate how the choice of the sample period affects the asymptotic prediction error covariance trace(P), where P satisfies (8), under the assumption that it is bounded. It is known since [30] that there exists a critical loss probability $\bar{\gamma}_c$ that rules the asymptotic behaviour of the error covariance: if $\bar{\gamma} < \bar{\gamma}_c$, P_k is bounded for all initial conditions and for each k, while it can diverge if $\bar{\gamma} \ge \bar{\gamma}_c$. In general, in the multidimensional case, it is not possible to compute $\bar{\gamma}_c$ analytically, but it admits upper and lower bounds. Many works, see e.g. [17,18,21], address the closed-form computation of $\bar{\gamma}_c$ in some particular case, for example when C is invertible. In the scalar case, $\bar{\gamma}_c = 1/\bar{a}^2$. We also investigate how the sampling period affects the tail probability of the prediction error $P(|e| > \beta(T))$. Clearly, from a design perspective, it is reasonable to identify the confidence interval to which the error belongs with very high probability similar to the benchmark 3σ in the Gaussian scalar case (where σ is the standard deviation),

[2] This value is chosen such that the non-arithmetic condition on the distribution of $\log \bar{a}$ is satisfied.

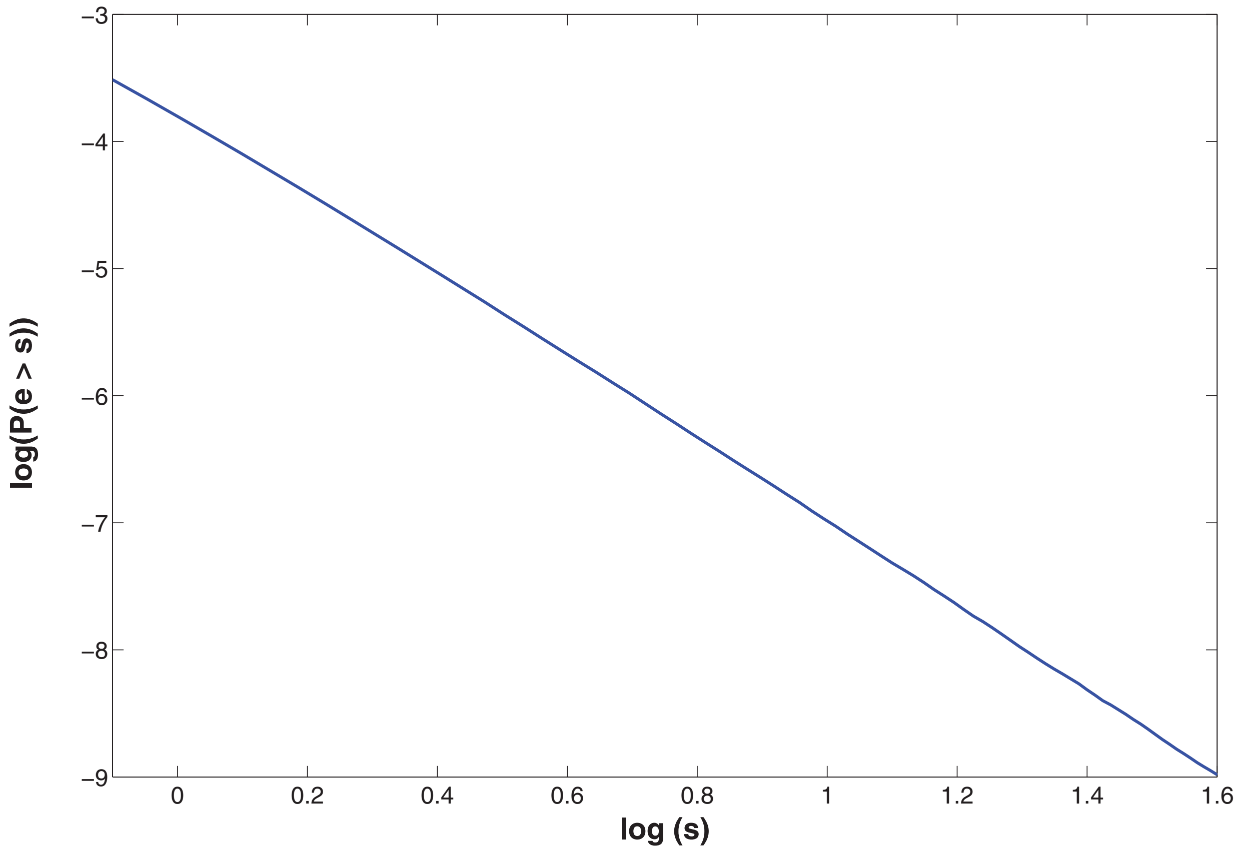

Fig. 3. Empirical log-log plot of the tail probability against the threshold.

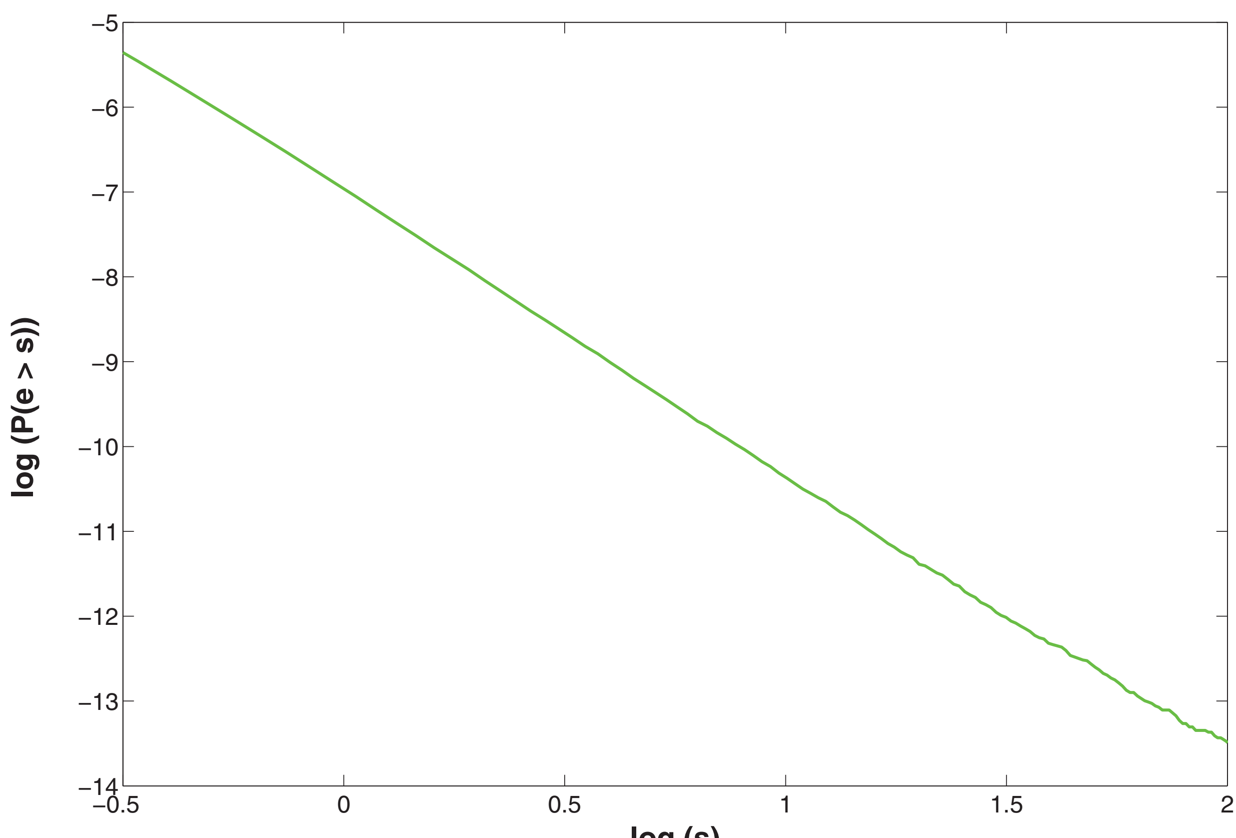

Fig. 4. Empirical log-log plot of the tail probability against the threshold (no measurement noise).

which gives a tail probability equal to 0.0027. Based on this observation, $\beta(T)$ is defined as the threshold for which the tail probability is $P(|e| > \beta(T)) = 0.0027$. We compare how the threshold $\beta(T)$ behaves with T, as compared to the threshold $3\sqrt{\text{trace}(P)}$, which would be the equivalent choice if the prediction error was strictly Gaussian. Since the prediction error stationary distribution has an asymptotic power-law tail behaviour, it is clear that $\beta(T)$ is expected to be larger than $3\sqrt{\text{trace}(P)}$. It is also of interest to observe the sampling period for which both $\beta(T)$ attains a minimum (say at T_1^*), and $3\sqrt{\text{trace}(P)}$ attains a minimum (say at T_2^*). It is in general difficult to obtain closed form expressions for T_1^* analytically. We therefore investigate this behaviour numerically through Monte Carlo simulations. This is also true for finding T_2^* in the general multivariable case. In the scalar, one can carry out a simple analysis to obtain the optimal sampling period T_2^* that minimizes $\bar{p}(T)$, which we discuss next. In order to proceed, one also needs to consider a specific form of $\bar{\gamma}$. To this end, we use a particular choice of dependence on T. We assume that every packet contains M number of bits, and the packet is lost even if a single bit is in error. This allows to write $\bar{\gamma} = 1 - (1 - BER(T))^M$, where $BER(T)$ denotes the bit error probability. In general, bit error probability depends on the underlying modulation schemes, and often takes the form of $Q(\sqrt{\rho \cdot SNR \cdot T/M})$, where $Q(\cdot)$ denotes the tail probability of a standard Gaussian random variable, i.e., $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp\left(-\frac{u^2}{2}\right)du$. Here SNR denotes the channel signal to noise ratio and ρ denotes a constant depending on the modulation scheme. For simplicity however, we choose $BER(T) = e^{-\nu T}$, where ν depends on SNR and the modulation scheme. Note that such an exponentially decaying bit error rate represents an upper bound on most $BER(T)$ of the form given by the $Q(\cdot)$ function above. In particular, for the case of noncoherent Binary Frequency Shift Keying (BFSK) modulation, the bit error rate is given by exactly an exponentially decaying function as above [22].

6.1. Optimal sampling period for minimizing $\bar{p}(T)$ (scalar case)

We recall that the prediction error follows (7) with $\bar{a} = \exp(aT)$, $\sigma^2 = \frac{\exp(2aT)-1}{2a}$. We take $c = 1$ and $q = 1$. We start with (8). Differentiating both sides of this equation with respect to T, and setting the first derivative $\frac{d\bar{p}(T)}{dT} = 0$, we obtain:

$$\bar{p}^2(T)\left(2a\bar{\gamma} + \frac{d\bar{\gamma}}{dT}\right) + \bar{p}(T)(1 + 2a\sigma_v^2) + \sigma_v^2 = 0,$$

Comparing this with Eq. (8), one can obtain, after some simplification:

$$\frac{2a}{e^{2aT}-1}\left(1 - e^{2aT}\bar{\gamma}\right) = -\left(2a\bar{\gamma} + \frac{d\bar{\gamma}}{dT}\right).$$

Rearranging the above equation, we get:

$$2a\left[\frac{1-\bar{\gamma}}{e^{2aT}-1}\right] = -\frac{d\bar{\gamma}}{dT}. \tag{14}$$

Replacing $\bar{\gamma} = 1 - (1 - BER(T))^M$ in (14), we obtain:

$$\frac{M\frac{dBER(T)}{dT}}{1 - BER(T)} = -\frac{2a}{e^{2aT}-1}.$$

Substituting $BER(T) = e^{-\nu T}$, after some algebraic manipulations, we obtain the following optimality equation for T_2^* that minimizes $\bar{p}(T)$:

$$T = \frac{1}{\nu}\log\left(1 + \frac{M\nu}{2a}\left(e^{2aT} - 1\right)\right). \tag{15}$$

Since we require $\bar{\gamma} < e^{-2aT}$ for $\bar{p}(T)$ to be finite, we can show that this implies $\log(1 - e^{-\nu T})^M > \log(1 - e^{-2aT})$. Since $M \geq 1$, we obtain a necessary condition on ν as $\nu > 2a$ for second moment stability.

It is clear that as $T \to 0$, the packet loss probability increases, driving $\bar{p}(T)$ to infinity, whereas when $T \to \infty$, $\bar{p}(T)$ also increases as the system becomes more unstable and the process noise variance σ^2 also increases. Therefore, $\bar{p}(T)$ must attain a minimum between $0 < T < \infty$. Clearly, the minimizing sampling period T_2^* must also satisfy the first order optimality condition (15). Ignoring the trivial solution $T = 0$ to (15), it is easy to show that there is only one solution $T_2^* > 0$ to (15), thus proving the uniqueness of the minimum. This follows from the fact that the function $h(T) = T - \frac{1}{\nu}\log\left(1 + \frac{M\nu}{2a}\left(e^{2aT} - 1\right)\right)$ has a negative derivative at $T = 0$, but its derivative becomes positive after $T = \frac{1}{2a}\log\left(\frac{1-\frac{2a}{M\nu}}{1-\frac{2a}{\nu}}\right)$, and remains positive, meaning that there is only one point where $h(T) = 0$ for $T > \frac{1}{2a}\log\left(\frac{1-\frac{2a}{M\nu}}{1-\frac{2a}{\nu}}\right)$.

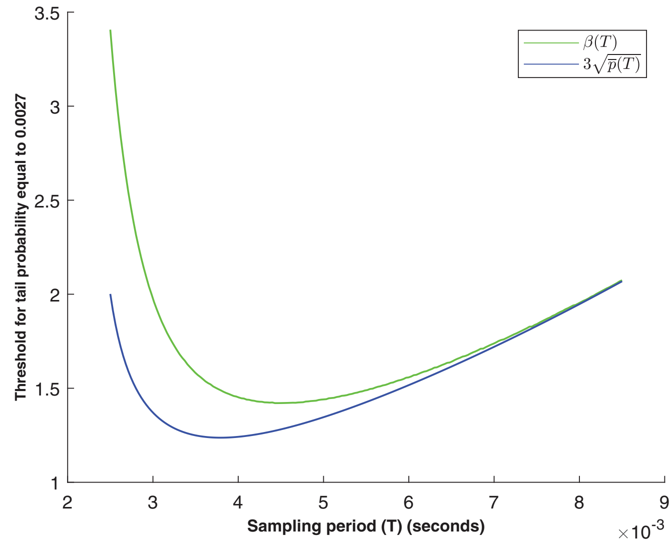

Fig. 5. Comparison of $\beta(T)$ and $3\sqrt{\bar{p}(T)}$ as a function of T, where $\beta(T)$ is such that $P(|e| > \beta(T)) = 0.0027$.

6.2. Numerical example: scalar case

The following Fig. 5 compares the two thresholds $\beta(T)$ and $3\sqrt{\bar{p}(T)}$, with $a = 100$, $\sigma_v^2 = 0.25$, where the bit error probability $BER(T)$ is given by $Q(\sqrt{\rho \cdot SNR \cdot T/M})$, with $\rho = 2$ (binary-phase-shift-keying modulation scheme). For our calculations, we assume that $M = 16$ and SNR=4.3 dB. The sampling time is varied within a range such that the minimum sampling time guarantees the packet loss probability stability threshold $\bar{\gamma}(T) < 1/\bar{a}^2(T)$. It is seen clearly that $\beta(T)$, the threshold based on the heavy tail distribution, is always greater than $3\sqrt{\bar{p}(T)}$ which is based on a Gaussian approximation. It is also seen that $T_1^* \approx 1.9369$ ms where as $T_2^* \approx 1.6354$ ms. This implies that due to the heavy-tail behaviour of the prediction error stationary distribution, one needs to use a higher sampling period (lower sampling rate) than would be recommended by a Gaussian assumption on the same distribution, if one is interested in ensuring a tail probability bound below a certain threshold. It is also noteworthy that as T decreases below or increases above T_1^* (T_2^*), the threshold $\beta(T)$ ($3\sqrt{\bar{p}(T)}$) increases as well, and the two thresholds almost approach each other when the sampling period becomes quite large. This is due to the fact that when the sampling period is approaching its lower limit (beyond which the stability threshold is violated), the packet loss probability is increasing - leading to a heavier tail dominated by the packet loss probability and increasing $\beta(T)$, at a higher rate than $3\sqrt{\bar{p}(T)}$. Similarly, when the sampling period $T\to\infty$, $\bar{\gamma}\to 0$, whereas the system becomes more unstable, along with increasing process noise variance σ^2. In this case the Gaussian approximation becomes a better fit with increasing sampling period. Since an associated LQ control cost can be computed in terms of the expected estimation error variance $\bar{p}(T)$, a similar behaviour (as $3\sqrt{\bar{p}(T)}$ in Fig. 5) with respect to the sampling period for an LQ control cost for a sampled data system controlled over an IEEE 802.15.4 based wireless local area network was observed in [19]. However, the behaviour of the estimation error tail probability was not analyzed either numerically or theoretically in [19].

6.3. Numerical example: multivariable case

In this section we consider a two-wheeled balancing robot, commonly referred to as *segway*, consisting of two wheels and a rigid body that must be kept upright. The rigid body contains two dc motors, the battery, and the electronic board. For simplicity, only longitudinal movements are allowed. The physical system is non-linear and so we decide to linearize it about the origin (in particular where the tilt angle is null). For the complete derivation of the model see [1]. The matrices of the continuous-time model are:

$$A=\begin{bmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 0 & 43.57 & -3.81 & 3.41\\ 0 & 55.22 & 1.97 & -2.25\end{bmatrix}\quad B=\begin{bmatrix}0\\0\\4.92\\-3.25\end{bmatrix}$$
$$C=\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}.$$

The input is the supply voltage of the dc motors, the state consists of the wheel angle, the tilt angle, and their derivatives, while the output is the wheel angle, that can be easily related to the linear position of the robot. The continuous eigenvalues are $\{0,\ 7.1184,\ -8.1501,\ -5.1372\}$. For what concerns the (continuous) process noise and the (discrete) measurement noise, the covariance matrices are:

$$\sqrt{Q}=\begin{bmatrix}10^{-2}\\10^{-3}\\0\\0\end{bmatrix}\qquad Q=\sqrt{Q}\sqrt{Q}^T\qquad R=10^{-5}.$$

For this simulation, the the bit error probability $BER(T)$ is given by $BER(T)=e^{-\nu T}$ with $\nu=750$ and the packet length is fixed to $M=64$. With this choice of the parameters, for the same sampling period, the packet loss probability is greater than the one of the previous example. As before, the sampling periods are such that the corresponding packet loss probability is below the critical value $\bar{\gamma}_c$ that guarantees the boundedness of the second order moment. As stated in [17], if $(\bar{A}, C)$ is detectable, $\bar{A}$ is diagonalizable, and the unstable eigenvalues of $\bar{A}$ have distinct magnitudes, the critical loss probability is $\bar{\gamma}_c=1/\max(|\lambda_i|^2)$, where λ_i, $i=1,\dots,n$, are the eigenvalues of $\bar{A}$. It is easy to prove that the discrete-time linearized model of the segway satisfies these hypotheses, so it is possible to find the critical loss probability for each T. It follows that the limit sampling period can be found from $\bar{\gamma}(T)<1/e^{2\cdot 7.1184\cdot T}$. In Fig. 6 we can see that the behaviour of the two thresholds is similar to the scalar case.

As expected, the threshold $\beta(T)$ such that $|e|>\beta(T)$ has a probability equal to 0.0027 is always greater than $3\sqrt{\text{trace}(P)}$, which would be the value of the threshold $\beta(T)$ if e was Gaussian. This fact confirms that the tail probability of the prediction error with packet loss is larger than the tail probability of the Gaussian case, i.e. confirms the heavy-tail behaviour of the prediction error for multivariable strictly unstable systems. Differently from the scalar example, the two thresholds are closer, i.e. the heavy-tail behaviour for the segway is less evident, due to the fact that the unstable eigenvalues of the two systems differ by a order of magnitude. This confirms that the unstable eigenvalues indicate how much the error distribution is heavy-tailed. The difference between the two thresholds is clearer when the packet loss is greater, that corresponds to the case when the sampling period is smaller. When packet loss probability decreases, i.e. when the sampling period increases, Gaussian distribution becomes a good approximation, since the curves are very close. These considerations validate the thesis that the packet loss is responsible for the heavy-tail distribution of the prediction error, in contrast with the case of no packet loss, where it has a Gaussian distribution. It is possible to find the optimal sampling periods, which are $T_1^*\approx 8$ ms for the heavy-tail case and $T_2^*\approx 7.5$ ms for the Gaussian case. These sampling periods minimize the thresholds, i.e. they ensure that the prediction error belongs (with the high probability $1-0.0027$) to the smallest confidence interval. This means that they are the most robust choice in order to avoid high deviations. Since $T_1^*>T_2^*$, in the case with packet loss one should use a higher sampling period than the one recommended by a Gaussian assumption, in order to have a tail probability bounded below a certain threshold. As expected, approaching the limit sampling period, both thresholds increase and tend to diverge, because the packet loss probability is closer to the critical value. In the same way, when the sampling period increases above the optimal points, the thresholds increase because the system becomes more unstable and the noise covariance $\bar{Q}$ becomes larger. These simulations suggest that, due to the heavy-tail behaviour, even when the condition for bounded error covariance is satisfied, high packet loss probability could return large deviations of the prediction error. If the prediction is used for control purpose, these not so rare large errors (with respect to the Gaussian case) can affect the control performance, resulting in unexpected practical consequences if only the error covariance matrix was considered as a performance metric in the design procedure. In particular, with unstable systems, the second order moment stability may not be sufficient to achieve the reliability required in the control applications.

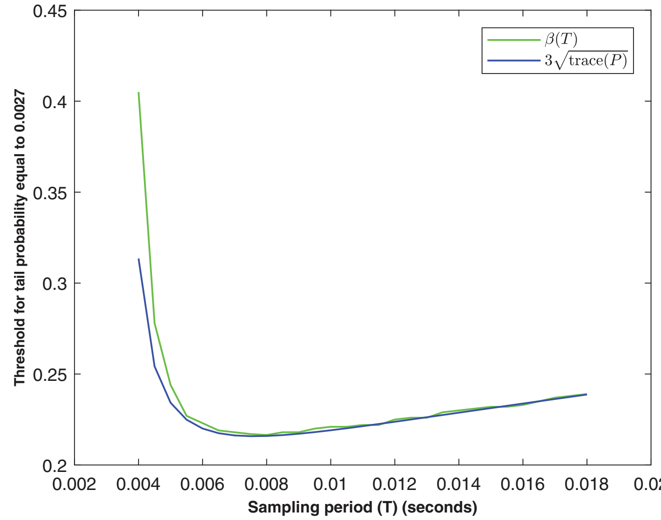

Fig. 6. Comparison of $\beta(T)$ and $3\sqrt{\text{trace}(P)}$ as a function of T, where $\beta(T)$ is such that $P(|e|>\beta(T))=0.0027$.

7. Conclusions

In this work we have studied sufficient conditions for the existence of a steady-state distribution for the prediction/estimation error in Kalman filtering with constant gains in the presence of packet losses. We have shown that such conditions are milder than the conditions for bounded error covariance. Moreover, we proved that for any linear filter, (i.e. not only in Kalman filtering), the steady-state distribution must be heavy-tailed if the system dynamics is strictly unstable, and if additional conditions exist such a distribution has a power-law tail whose exponent depends on the packet loss probability. This implies that confidence bounds obtained by computing error covariance and assuming a Gaussian distribution are in general optimistic, i.e. the probability of having large deviations in estimation errors are larger than expected. Nonetheless, we numerically illustrated that for realistic unstable systems, such a difference can be very small in the range of optimal operating conditions in terms of transmission rate. Future works will explore the extension of these results to Markovian packet dropouts and optimal rate selection using more realistic communication protocols.

Acknowledgement

This work was supported by the Swedish Research Council [grant Dnr 621-2013-5395] and by University of Padova project MAgIC [SCHE_SID17_01].

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejcon.2019.05.003.

References

[1] R. Antonello, L. Schenato, Laboratory activity 4: Longitudinal statespace control of the balancing robot, Technical report, University of Padova, Department of Information Engineering, 2017 http://automatica.dei.unipd.it/tl_files/utenti/lucaschenato/SEGWAY_GUIDE.pdf.
[2] A. Brandt, The stochastic equation $y_{n+1} = a_n y_n + b_n$ with stationary coefficients, Adv. Appl. Prob. 18 (1986) 211–220.
[3] D. Buraczewski, E. Damek, T. Mikosch, Stochastic Models with Power-Law Tails: The Equation $X = AX + B$, Springer International Publishing, 2016.
[4] A. Censi, Kalman filtering with intermittent observations: Convergence for semi-Markov chains and an intrinsic performance measure, IEEE Trans. Autom. Control 56 (2) (2010) 376–381.
[5] S. Dey, L. Schenato, Heavy-tails in Kalman filtering with packet losses: confidence bounds vs. second moment stability, in: Proceedings of the European Control Conference, 2008.
[6] C. Goldie, Implicit renewal theory and tails of solutions of random equations, Ann. Appl. Probab. 1 (1) (1991) 126–166.
[7] C. Goldie, R. Grubel, Perpetuities with thin tails, Adv. Appl. Prob. 28 (1996) 463–480.
[8] I. Goldsheid, G. Margulis, Lyapunov indices of a product of random matrices, Russian Math. Surv. 44 (1989) 11–71.
[9] A. Grincevicius, One limit distribution for a random walk on the line, Lith. Math. J. 15 (1975) 580–589.
[10] J. Hespanha, P. Naghshtabrizi, Y. Xu, A survey of recent results in networked control systems, Proc. IEEE 95 (1) (2007) 138–162.
[11] P. Hitczenko, J. Wesolowski, Perpetuities with thin tails, revised, Ann. Appl. Probab. 19 (2009) 2080–2101.
[12] M. Huang, S. Dey, Stability of Kalman filtering with Markovian packet losses, Automatica 43 (2007) 598–607.
[13] S. Kar, B. Sinopoli, J. Moura, Kalman filtering with intermittent observations: weak convergence to a stationary distribution, IEEE Trans. Autom. Control 57 (2012) 405–420.
[14] H. Kesten, Random difference equations and renewal theory for products of random matrices, Acta Math. 131 (1973) 207–248.
[15] M. Loeve, Probability Theory, 2nd ed., Van Nostrand, New York, 1960.
[16] D. Mannion, Products of 2×2 random matrices, Ann. Appl. Probab. 3 (4) (1993) 1189–1218.
[17] Y. Mo, B. Sinopoli, A characterization of the critical value for Kalman filtering with intermittent observations, in: Proceedings of the 47th IEEE Conference on Decision and Control, IEEE, 2008, pp. 2692–2697.
[18] Y. Mo, B. Sinopoli, Kalman filtering with intermittent observations: Tail distribution and critical value, IEEE Trans. Auto. Control 57 (3) (2012) 677–689.
[19] P. Park, J. Araujo, K. Johansson, Wireless networked control system co-design, in: Proceedings of the International Conference on Networking, Sensing and Control, Delft, the Netherlands, 2011, pp. 486–491.
[20] S.Y. Park, A. Sahai, Intermittent Kalman filtering: Eigenvalue cycles and nonuniform sampling, in: Proceedings of the American Control Conference, 2011.
[21] K. Plarre, F. Bullo, On kalman filtering for detectable systems with intermittent observations, IEEE Trans. Autom. Control 54 (2) (2009) 386–390.
[22] J. Proakis, M. Salehi, Digital Communications, 5th ed., McGraw-Hill Education, 2007.
[23] V. Protasov, R. Jungers, Lower and upper bounds for the largest Lyapunov exponent of matrices, Linear Algebra Appl. 438 (2013) 4448–4468.
[24] E.R. Rohr, D. Marelli, M. Fu, Kalman filtering with intermittent observations: On the boundedness of the expected error covariance, IEEE Trans. Autom. Control 59 (2010) 2734–2738.
[25] A. Sahai, S. Mitter, The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication. part I: scalar systems, IEEE Trans. Inf. Theory 52 (8) (2006) 3369–3395.
[26] B. de Saporta, Tail of the stationary solution of the stochastic equation $y_{n+1} = a_n y_n + b_n$ with Markovian coefficients, Stoch. Process. Appl. 115 (2005) 1954–1978.
[27] L. Schenato, Kalman filtering for networked control systems with random delay and packet loss, IEEE Trans. Autom. Control 53 (2008) 1311–1317.
[28] L. Shi, M. Epstein, R. Murray, Kalman filtering over a packet-dropping network: a probabilistic perspective, IEEE Trans. Autom. Control 55 (3) (2010) 594–604.
[29] L. Shi, M. Epstein, A. Tiwari, R. Murray, Estimation with information loss: asymptotic analysis and error bounds, in: Proceedings of the 44th IEEE CDC, Seville, Spain, 2005, pp. 1215–1221.
[30] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, S. Sastry, Kalman filtering with intermittent observations, IEEE Trans. Autom. Control 49 (9) (2004) 1453–1464.
[31] J. Wu, L. Shi, L. Xie, K. Johansson, An improved stability condition for Kalman filtering with bounded Markovian packet losses, Automatica 62 (2015) 32–38.
[32] K. You, M. Fu, L. Xie, Stability of Kalman filtering with Markovian packet losses, Automatica 47 (12) (2011) 2647–2657.
[33] K. You, L. Xie, Survey of recent progress in networked control systems, Acta Sinica 39 (2) (2013) 101–117.