Heavy-tails in Kalman filtering with packet losses: confidence bounds vs second moment stability

S. Dey *Senior Member, IEEE,*, and L. Schenato *Fellow, IEEE*

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 scalar 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. By drawing results from Renewal Theory, in this work we show that under the assumption that packet loss probability is smaller than unity, and the system is on average contractive, a stationary distribution always exists and is heavy-tailed, i.e. its absolute moments beyond a certain order do not exist. We also show that under additional technical assumptions, 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 exponential decay of packet loss probability with respect to the sampling period. In order to minimize the expected value of 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.

Index Terms—Kalman Filtering, Packet Losses, Heavy-tailed Distributions, Power-Law Tails

I. INTRODUCTION

Estimation and control in the presence of unreliable communication has drawn the interest of a large body of researches in the past decade as a result of convergence of communication and control due to ubiquitous presence of wireless communication [1], [2]. In particular, much attention has been placed on the problem of optimal estimation in the presence of packet losses via Kalman Filtering since [3]. 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. The first major result was to show the existence of a critical packet loss probability for the boundedness of the second moment under an i.i.d. packet loss scenario [3]. Later this analysis has been extended to Markov packet losses [4], [5], [6], to the computations of upper and lower bounds for such critical packet loss probability [7], [8], [6], to the existence and type of distribution for the error covariance [9], [10]. 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 [11], eigenvalue cycles [12] and non-degeneracy [13].

In this work we concentrate on a continuous-time stochastic *unstable scalar linear system* which is sampled with sampling period T. Although most of the recent results have rightfully concentrated on the more general multivariable systems, specifically on the analysis of the error covariance matrix and its distribution of the distribution of some functions of it as the trace, we believe that not enough attention has been directed towards understanding the actual distribution of the prediction/estimation error. More specifically, by addressing this problem, we show that a steady state distribution exists under much milder conditions on the packet loss probability. Indeed for a scalar unstable system such distribution exists as soon as the packet loss probability in not equal to unity. In other words, there are scenarios in which the second moment does not exist, but the probability that the estimation error is within a specified interval is bounded. Moreover, by exploiting results developed by the stochastic systems community in the area of Renewal Theory and Random Difference Equations [14], [15], [16], [17], [18] since the 70's, it is possible to explicitly characterize the tail distribution of the Kalman filter of the estimation error. More specifically, we can show that such distribution is heavy-tailed and under some technical condition such as *non-arithmetic* support of 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 [19], which however is 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 some conditions mentioned above for the power-law tail for the estimation error distribution, i.e. we recover the same results observed previously in [13]. 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 [4]. Another major contribution of our work is that, since the error distribution is heavy-tailed even when the system is second-moment stable, this implies that the confidence bounds can be rather different from what one would obtain by 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 [3]. In other words, the $3\sqrt{\bar{p}}$ estimation of the confidence bounds, where $\bar{p}$ is the expected 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 a stochastic continuous-time scalar unstable system both in terms of minimizing the expected second moment $\bar{p}$ and in terms of the confidence bound for the steady state error with a 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.

II. SYSTEM MODEL

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

$$dx(t) = ax(t)dt + u(t)dt + d\bar{w}(t) \qquad (1)$$

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

$$x_{k+1} = \bar{a}(T)x_k + b(T)u_k + w_k \qquad (2)$$

where $\bar{a}(T) = e^{aT}$, $b(T) = \frac{e^{aT}-1}{a}$ and w_k is independent and identically Gaussian distributed with zero mean and variance $\sigma^2(T) = \frac{e^{2aT}-1}{2a}$. We assume the discrete-time measurements of this system are also scalar and given by a linear equation:

$$y_k = x_k + v_k \qquad (3)$$

where v_k is the measurement noise, independent of w_k, and x_0, and is also independent and identically Gaussian distributed with zero mean and variance σ_v^2. In the subsequent analysis, we focus on unstable systems only such that $a > 0$ or $\bar{a}(T) > 1$. We also assume that $E[|x_0|^r] < \infty, r \geq 1$. These measurements are transmitted to a remote estimator over a lossy channel such that each measurement is either received according to a Bernoulli process $\gamma_k \in \{0,1\}$ (independent of v_k, w_k, x_0), such that the $\gamma_k = 0$ (loss of packet) with probability $\bar{\gamma}(T)$. 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 $\mathcal{Z}_k \triangleq \{y_0, y_1, \ldots y_k, \gamma_0, \gamma_1, \ldots, \gamma_k\}$, the remote estimator produces a constant gain Kalman predictor

$$\hat{x}_{k+1|k} = \bar{a}(T)\hat{x}_{k|k-1} + b(T)u_k + K_p\gamma_k(y_k - \hat{x}_{k|k-1}) \qquad (4)$$

where K_p is a constant Kalman prediction gain chosen appro-

priately. Define the prediction error $e_k = x_k - \hat{x}_{k|k-1}$. Note that the above Kalman predictor is suboptimal in the sense that it does not use the optimal time-varying Kalman gain as in [3]. This simplification is used here to make the subsequent analysis simpler. A suitable choice of K_p is discussed in [20] for example, where it was shown that under the condition that $\lim_{k\to\infty} E(e_k^2) < \infty$ (i.e., the packet loss probability is less than $1/\bar{a}(T)^2$), K_p can be chosen to be the gain that minimizes $\lim_{k\to\infty} E(e_k^2)$, and this choice results in a marginal performance loss in terms of expected estimation error variance. Indeed, $K_p = \frac{\bar{a}(T)\bar{p}(T)}{\bar{p}(T)+\sigma_v^2}$, where $\bar{p}(T) \triangleq \lim_{k\to\infty} E(e_k^2)$ satisfies the modified algebraic Riccati equation:

$$\bar{p}(T) = \bar{a}(T)^2\bar{p}(T) + \sigma^2(T) - (1-\bar{\gamma}(T))\frac{\bar{a}(T)^2\bar{p}(T)^2}{\bar{p}(T)+\sigma_v^2} \qquad (5)$$

Clearly, it is seen that $K_p < \bar{a}(T)$, and it can also be shown from (5), that $(\bar{a}(T) - K_p) < 1$.

With this choice of the Kalman prediction gain, the recursion for the prediction error $e_k = x_k - \hat{x}_{k|k-1}$ is then given by

$$e_{k+1} = (\bar{a}(T) - K_p\gamma_k)e_k + (w_k - K_p\gamma_k v_k) \qquad (6)$$

In the next section, we first provide a simple proof that e_k is asymptotically heavy-tailed in that its absolute moments higher than a certain order go to infinity as $k \to \infty$.

Before we proceed, we introduce a few notations that will be used in the next section. 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. $\log^+ x = \log x$ for $x \geq 1$ and 0 for $x \in (0,1)$.

III. HEAVY TAIL PROPERTIES OF PREDICTION ERROR

Asymptotically infinite absolute moments of the prediction error

In the first instance, we work with the prediction error equation (6) when there is no measurement noise, that is, $v_k = 0$ and consequently $K_p = \bar{a}(T)$. The rationale is to keep the proof simple keeping in mind the fact that the absolute moments of the prediction error with measurement noise are bigger than those without the measurement noise. Hence if an absolute moment of the prediction error is infinite as $k \to \infty$ in the absence of measurement noise, it will be also infinite in the presence of measurement noise. A more rigorous statement is given at the end of the following proof.

First we quote the following moment inequality (see [21] p. 263 for a proof):

Lemma 1: If $E[|X|^r] < \infty$, $E[|Y|^r] < \infty$ and $E[Y|X] = 0$ almost sure, then

$$E[|X + Y|^r] \geq E[|X|^r], \forall r \geq 1 \qquad (7)$$

Now specialising (6) in the no measurement noise case, we have

$$e_{k+1} = \bar{a}(T)(1 - \gamma_k)e_k + w_k \qquad (8)$$

Therefore defining $X = \bar{a}(T)(1-\gamma_k)e_k$, and $Y = w_k$, we see that for $r \geq 1$, $E[|X|^r] = \bar{a}(T)^r\bar{\gamma}(T)E[|e_k|^r]$, which is finite for some $k \leq k_0$, and $E[|w_k|^r] < \infty$ from Gaussianity, as well as $E[w_k|X] = E[w_k] = 0$ from the modelling assumptions. Therefore we have $E[|e_{k+1}|^r] \geq \bar{a}(T)^r\bar{\gamma}(T)E[|e_k|^r]$ for $k \geq$

k_0. Since $\bar{a}(T) > 1$, it automatically follows that $\bar{a}(T)^r \bar{\gamma}(T) > 1$ for some $r > \bar{r} \geq 1$ and hence $E[|e_{k+1}|^r] > E[|e_k|^r]$ for $r \geq \bar{r}$, and hence $\lim_{k \to \infty} E[|e_{k+1}|^r] \to \infty$ for $r \geq \bar{r}$. This gives us the desired result that the absolute moments above a certain order of the prediction error are asymptotically heavy-tailed. By applying Lemma 1 to the recurrence equation (6) involving measurement noise, it is clear that for $r \geq \bar{r}$, $\lim_{k \to \infty} E[|e_{k+1}|^r] \to \infty$ also in the measurement noise case. We summarize this result in the following Proposition.

Proposition 1: There exists a $\bar{r} \geq 1$ such that for $r \geq \bar{r}$, $\lim_{k \to \infty} E[|e_{k+1}|^r] \to \infty$ where e_k is the prediction error follows the equation (6), thus exhibiting a heavy-tail property.

A similar result holds for the corresponding Kalman filtering error.

Existence of stationary solution to the prediction error

It is well known that for a random stochastic difference equation $Y_{k+1} = a_k Y_k + b_k$, where (a_k, b_k) is stationary and ergodic along with $E[\log|a_0|] < 0$ and $E \log^+ |b_0| < \infty$, $Y_n, n \in \mathbb{Z}$ has a unique stationary solution given by $Y_n = \sum_{k=0}^{\infty} a_{n-1} a_{n-2} \ldots a_{n-k} b_{n-1-k}$ [16], [22]. Earlier versions of these types of stationarity results can also be found for the special case of (a_k, b_k) being independent and identically distributed (*i.i.d*) exist in many works, see Theorem 2.1.3 in [18] for example. This result is especially relevant to our case, as we consider Bernoulli packet drops.

Now we specifically consider (6) and rewrite it as $e_{k+1} = f_k e_k + z_k$, where $f_k = (a(T) - K_p \gamma_k), z_k = (w_k - K_p \gamma_k v_k)$. We can show that (since $K_p < \bar{a}(T)$, and $\bar{a}(T) > 1$) $P(f_k = 0) = 0$. Also, since w_k, v_k are both zero mean Gaussian, $E[\log^+ |z_k|] < \infty$. We make the following additional assumption:

Assumption 1:

$$E[\log|f_k|] = \gamma(T) \log(\bar{a}(T))$$
$$+ (1 - \gamma(T)) \log((\bar{a}(T) - K_p) < 0 \quad (9)$$

Assumption 1 is sufficient to guarantee the existence of a stationary solution to (6) and is known as the contractive case [18]. Under this assumption, let us denote the stationary distribution of e_k in (6) as $g_\infty(e)$. Then $g_\infty(e)$ satisfies the following integral equation:

$$g_\infty(z) = \bar{\gamma}(T) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2(T)}} e^{-\frac{(z - \bar{a}(T)e)^2}{2\sigma^2(T)}} g_\infty(e) de$$
$$+ (1 - \bar{\gamma}(T)) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\delta^2(T)}} e^{-\frac{(z - (\bar{a}(T) - K_p)e)^2}{2\delta^2(T)}} g_\infty(e) de$$
$$(10)$$

where $\delta^2(T) = \sigma^2(T) + K_p^2 \sigma_v^2$ is the variance of $w_k - K_p v_k$ in (6). 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 (10) 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 leads to a linear equation involving $\bar{g}$, which can be solved iteratively until the solution converges to within a given tolerance. As always, the larger the number of discretization points (the longer the vector $\bar{g}$), the better the approximation. Figure 1 below compares the stationary distribution of the prediction error for a system with $\bar{a}(T) = 1.9251, K_p = 1.4164, \sigma_v^2 = 0.25, \sigma^2(T) = 0.6960$, corresponding to a sampling period of $T = 6.55$ milliseconds. The blue graph shows the density obtained using the discretized approximation to the integral equation (10) 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.
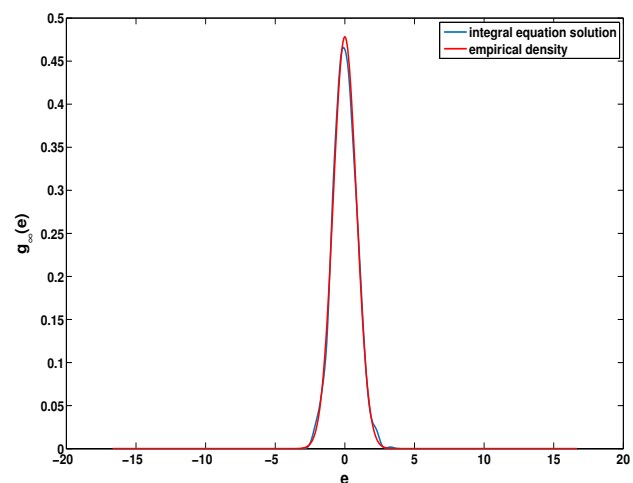

Fig. 1. Plot comparing the stationary distribution of the prediction error obtained via approximate solution to the integral equation (blue) and en empirical method (red)

Asymptotic power-law tails of the prediction error

For the purpose of this section, we again rewrite (6) as $e_{k+1} = f_k e_k + z_k$. 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 $E \stackrel{d}{=} FE + Z$, where $\stackrel{d}{=}$ denotes equality in distribution, similar to [17], [18]. We now assume that (F, Z) is a general i.i.d. random process (of course, our particular case satisfies this assumption as discussed above). 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 $s \to \infty$, $P(E > s) \sim C_1 s^{-\alpha}$ 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 [17], 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 [18], as Theorem 2.4.4.

We now quote Theorem 2.4.4 from [18], using the notations of our paper.

Theorem 1: [18] 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 $E[F^\alpha] = 1$, $E[|Z|^\alpha] < \infty$, and $E[F^\alpha \log^+ F] < \infty$. 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

$$P(E > s) \sim c_+ s^{-\alpha}, \ P(E < -s) \sim c_- s^{-\alpha}$$
$$c_+ = \frac{E\left[(FE + Z)_+^\alpha - (FE)_+^\alpha\right]}{\alpha m_\alpha}$$
$$c_- = \frac{E\left[(FE + Z)_-^\alpha - (FE)_-^\alpha\right]}{\alpha m_\alpha} \quad (11)$$

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

Before we discuss the implications of this result in our specific case of (6), we note that under the assumptions (i), (ii) and (iii) of Theorem 1 above, it automatically follows that (i) $-\infty \leq E[\log F] < 0$ (thus guaranteeing stationarity) and (ii) $0 < m_\alpha < \infty$ - see Lemma 2.2 of [17]. A proof of the above theorem using renewal theory is provided also in [18].

Now let us apply Theorem 1 to (6) rewritten as $e_{k+1} = f_k e_k + z_k$. It is clear that $f_k \in \{\bar{a}(T), \bar{a}(T) - K_p\}$ and therefore $f_k > 0$, and since all absolute moments of z_k are finite. Finally, we also have that, since the stationary distribution of e_k is continuous, assumption (iv) of the above theorem is also satisfied. The non-arithmetic requirement on the distribution of $\log(f_k)$ essentially means that $\frac{\log(\bar{a}(T))}{\log(\bar{a}(T) - K_p)}$ cannot be a rational number. We will discuss the case when the distribution of $\log(f_k)$ is arithmetic a little later. We also note that for any $\bar{\gamma}(T) < 1$, one can show (since $0 < \bar{a}(T) - K_p < 1$, and $\bar{a}(T) > 1$) that there exists an $\alpha > 0$ such that the condition $E[f_k^\alpha] = 1$ is satisfied. Finally, it is easy to check that $E[f_k^\alpha \log^+ f_k] = \bar{a}(T)^\alpha \log(\bar{a}(T)) \bar{\gamma}(T) < \infty$. Therefore all the assumptions of the previous theorem apply and we can conclude that $g_\infty(e)$ satisfies the asymptotic power law tail behaviour as described in Theorem 1. 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 (11) are equal and positive.

The case where the distribution of $\log f_k$ conditioned on $f_k \neq 0$ is *arithmetic*, the analysis of the tail probability is more complex and was carried out in [15], [23]. The basic result is that in this case one can prove [18] that there are constants $c_a > c_b > 0$ such that

$$c_a \leq \liminf_{s \to \infty} s^\alpha P(|E| > s) \leq \limsup_{s \to \infty} s^\alpha P(|E| > s) \leq c_b$$
$$(12)$$

Remark 1: If one considers the no measurement noise case, where the prediction error recursion follows (8), then it is seen that in this case $f_k = \bar{a}(T)$ with probability $\bar{\gamma}(T)$ and 0 with probability $(1 - \bar{\gamma}(T))$. 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 [15], [23]. 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 (12) that has proved to be difficult.

Remark 2: The power law exponent α in our context is the solution to the equation $\bar{a}(T)^\alpha \bar{\gamma}(T) = 1$, i.e.

$$\alpha = \frac{\log\left(\frac{1}{\bar{\gamma}(T)}\right)}{\log \bar{a}(T)} \quad (13)$$

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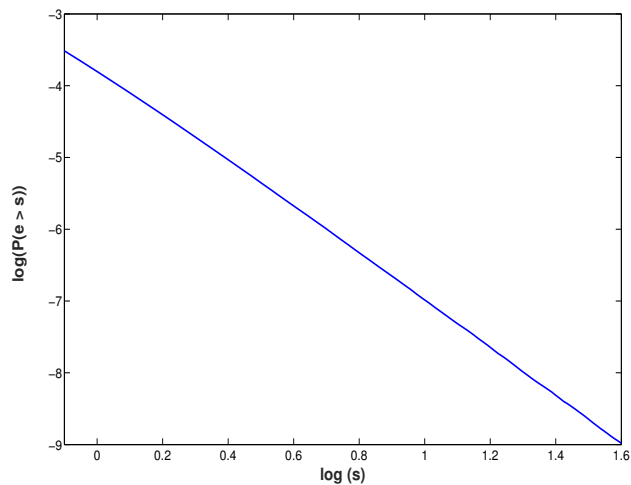

Fig. 2. Empirical log-log plot of the tail probability against the threshold

Figure 2 illustrates a plot of the $\log(P(e > s))$ versus $\log(s)$

---

[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.

for the following parameters: $\bar{a}(T) = e^{\sqrt{\frac{1}{17}}}$,[2] $K_p = 0.7$, $\sigma^2(T) = 0.01$, $\sigma_v^2 = 0.1$, and $\bar{\gamma}(T) = 0.4$. 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 (11) is given by 0.0270.
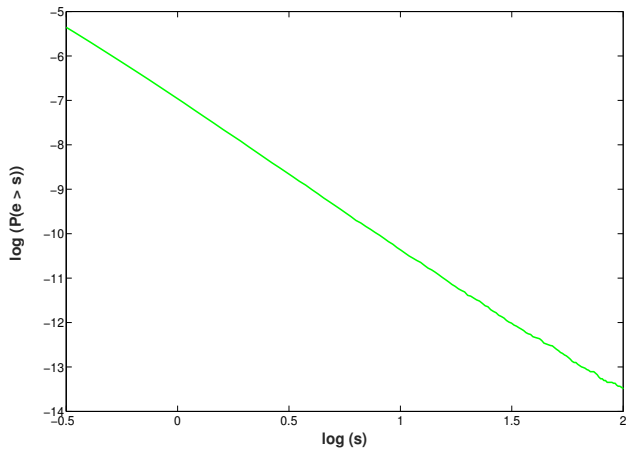

Fig. 3. Empirical log-log plot of the tail probability against the threshold (no measurement noise)

Similarly, Figure 3 illustrates a plot of $\log(P(e > s))$ versus $\log(s)$ in the no measurement noise case, with $\bar{a}(T) = 1.3$, $\sigma^2(T) = 0.01$ and $\bar{\gamma}(T) = 0.4$. 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 3: 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 [22]. However, the technical conditions under which these results hold are significantly more complicated, especially the conditions involving the non-arithmetic distributions. Extension of our results to the multivariable case is therefore left for future work.

IV. 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 [3] or tail bounds such as in [24]. In [13], 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 now investigate how the choice of the sample period affects the expected prediction error variance $\bar{p}(T)$ under the assumption that it is bounded (i.e. when $\bar{\gamma}(T) < 1/\bar{a}^2(T)$). We also investigate how the sampling period affects the tail probability of the prediction error: $P(|e| > \beta(T))$. Clearly, if from a design perspective, one wanted to guarantee a bound on the tail probability such

that $P(|e| > \beta) \leq 0.0027$, which is the Gaussian (with zero mean) tail probability beyond 3σ, where σ is the standard deviation. We compare how the threshold $\beta(T)$ behaves with T, as compared to the threshold $3\sqrt{\bar{p}(T)}$, 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 $3\sqrt{\bar{p}(T)}$. It is also of interest to observe the sampling period for which both $\beta(T)$ attains a minimum (say at T_1^*), and $3\sqrt{\bar{p}(T)}$ 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. On the other hand, one can carry out a simple analysis to obtain the optimal sampling period T_2^* that minimizes $\bar{p}(T)$, which we discuss next.

Optimal sampling period for minimizing $\bar{p}(T)$

We recall that the prediction error follows (6) with $\bar{a}(T) = \exp(aT)$, $\sigma^2(T) = \frac{\exp(2aT)-1}{2a}$. We start with (5). Differentiating both sides of this equation with respect to T, and setting the first derivative $\frac{d\bar{p}(T)}{dT} = 0$, we obtain

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

Comparing this with equation (5), one can obtain, after some simplification,

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

Rearranging the above equation, we get

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

In order to obtain a solution to this equation, one needs to consider a specific form of $\bar{\gamma}(T)$. 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

$$\bar{\gamma}(T) = 1 - (1 - BER(T))^M$$

where $BER(T)$ denotes the bit error probability. In general, bit error probability depends on the underlying modulation schemes, and often take the form of $Q(\sqrt{\rho SNRT/M})$, where $Q(.)$ 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 $Q(.)$ 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 [25].

Replacing $\bar{\gamma}(T) = 1 - (1 - BER(T))^M$ 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 $\bar{p}(T)$:

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

Since we require $\bar{\gamma}(T) < e^{-2aT}$ 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 $\sigma^2(T)$ 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 $T > \frac{1}{2a}\log\left(\frac{1 - \frac{2a}{L\nu}}{1 - \frac{2a}{beta}}\right)$, and remains positive, meaning that there is only one point where $h(T) = 0$ for $T > \frac{1}{2a}\log\left(\frac{1 - \frac{2a}{L\nu}}{1 - \frac{2a}{beta}}\right)$.

Numerical Example

The following Figure 4 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 $Q(\sqrt{\rho SNRT/M}$, 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 $T_1^* \approx 1.9369$ milliseconds where as $T_2^* \approx 1.6354$ milliseconds. 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 $T \to \infty$, $\bar{\gamma}(T) \to 0$, whereas the system becomes more unstable, along with increasing process noise variance $\sigma^2(T)$. 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 Figure 4) 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 [26]. However, the behaviour of the estimation error tail probability was not analyzed either numerically or theoretically in [26].
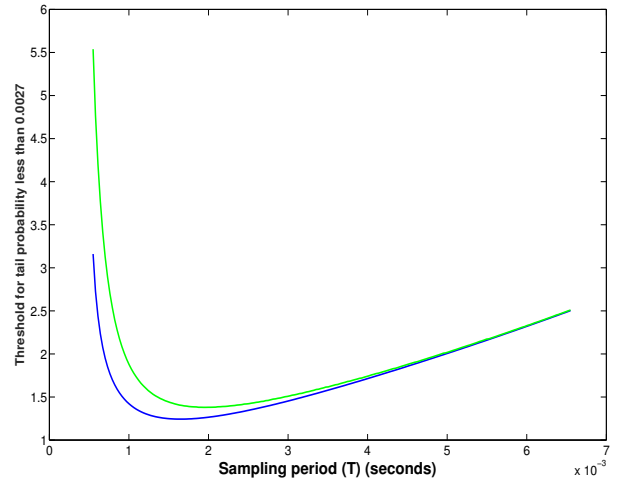

Fig. 4. Comparison of $\beta(T)$ and $3\sqrt{\bar{p}(T)}$ as a function of T, where $\beta(T)$ is such that $P(|e| > \beta(T)) \leq 0.0027$.

V. CONCLUSIONS

In this paper, we established the existence of a steady state distribution and its heavy tailed behaviour of the prediction/estimation error in Kalman filtering with packet losses for an unstable scalar linear system. The existence of the steady state distribution holds under milder conditions than previously known results for the existence of the steady state distribution of the estimation error variance conditioned on the packet loss process. Under additional conditions on the distribution of the state transition coefficient, it is shown that the prediction/estimation error steady state distribution has a heavy-tailed behaviour, and in particular, exhibits an asymptotic power law tail. Using results from Renewal Theory and Random Difference Equations, the exact behaviour of the tail probability can be characterized in terms of its power law exponent and the associated constant gain term. We also investigate the effect of the sampling period when a continuous-time stochastic unstable scalar system is sampled for measurements, which in turn are transmitted over a channel with packet loss for remote Kalman filtering. The effect of this sampling period is investigated in terms of both the expected error covariance and the estimation error tail probability, and it is seen the heavy-tail behaviour results in a larger sampling period while considering tail probabilities for a certain confidence bound, than that would be implied by a strictly Gaussian assumption on the estimation error distribution, with a variance equal to the expected error covariance. This finding should be also

useful for practitioners working with remote estimation/control of sampled-data continuous time unstable linear systems in the presence of packet loss. Future works will explore the extension of these results to the multivariable case along with Markovian packet dropouts, by carefully exploiting the results in [22].

REFERENCES

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

[2] K. You and L. Xie, "Survey of recent progress in networked control systems," *Acta Sinica*, vol. 39, no. 2, pp. 101–117, 2013.

[3] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, September 2004.

[4] M. Huang and S. Dey, "Stability of kalman filtering with markovian packet losses," *Automatica*, vol. 43, pp. 598–607, 2007.

[5] K. You, M. Fu, and L. Xie, "Stability of kalman filtering with markovian packet losses," *Automatica*, vol. 47, no. 12, pp. 2647–2657, 2011.

[6] J. Wu, L. Shi, L. Xie, and K. Johansson, "An improved stability condition for kalman filtering with bounded markovian packet losses," *Automatica*, vol. 62, pp. 32–38, 2015.

[7] L. Shi, M. Epstein, and R. Murray, "Kalman filtering over a packet-dropping network: a probabilistic perspective," *IEEE Transactions on Automatic Control*, vol. 55, no. 3, pp. 594–604, 2010.

[8] E. R. Rohr, D. Marelli, and M. Fu, "Kalman filtering with intermittent observations: On the boundedness of the expected error covariance," *IEEE Transactions on Automatic Control*, vol. 59, pp. 2734–2738, 2010.

[9] A. Censi, "Kalman filtering with intermittent observations: Convergence for semi-markov chains and an intrinsic performance measure," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 376 – 381, 2010.

[10] S. Kar, B. Sinopoli, and J. Moura, "Kalman filtering with intermittent observations: weak convergence to a stationary distribution," *IEEE Transactions on Automatic Control*, vol. 57, pp. 405–420, 2012.

[11] K. Plarre and F. Bullo, "On kalman filtering for detectable systems with intermittent observations," *IEEE Transactions on Automatic Control (to appear)*, vol. 54, no. 2, pp. 386 – 390, 2009.

[12] S. Y. Park and A. Sahai, "Intermittent kalman filtering: Eigenvalue cycles and nonuniform sampling," in *Proceedings of the 2011 American Control Conference*, 2011.

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

[14] H. Kesten, "Random difference equations and renewal theory for products of random matrices," *Acta Math.*, vol. 131, pp. 207–248, 1973.

[15] A. Grincevicius, "One limit distribution for a random walk on the line," *Lithuanian Math. Journal*, vol. 15, pp. 580–589, 1975.

[16] A. Brandt, "The stochastic equation $y_{n+1} = a_n y_n + b_n$ with stationary coefficients," *Adv. Appl. Prob.*, vol. 18, pp. 211–220, 1986.

[17] C. Goldie, "Implicit renewal theory and tails of solutions of random equations," *The Annals of Applied Probability*, vol. 1, no. 1, pp. 126–166, 1991.

[18] D. Buraczewski, E. Damek, and T. Mikosch, *Stochastic Models with Power-Law Tails: The equation $X = AX + B$*. Springer International Publishing, 2016.

[19] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication. part i: Scalar systems," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3369 – 3395, 2006.

[20] L. Schenato, "Kalman filtering for networked control systems with random delay and packet loss," *IEEE Transactions on Automatic Control*, vol. 53, pp. 1311–1317, 2008.

[21] M. Loeve, *Probability Theory, 2nd Edition*. New York: Van Nostrand, 1960.

[22] B. de Saporta, "Tail of the stationary solution of the stochastic equation $y_{n+1} = a_n y_n + b_n$ with Markovian coefficients," *Stochastic Processes and their Applications*, vol. 115, pp. 1954–1978, 2005.

[23] C. Goldie and R. Grubel, "Perpetuities with thin tails," *Adv. Appl. Prob.*, vol. 28, pp. 463–480, 1996.

[24] L. Shi, M. Epstein, A. Tiwari, and R. Murray, "Estimation with information loss: Asymptotic analysis and error bounds," in *Proc. of the 44th IEEE CDC*, Seville, Spain, 2005, pp. 1215–1221.

[25] J. Proakis and M. Salehi, *Digital Communications, 5th Edition*. McGraw-Hill Education, 2007.

[26] P. Park, J. Araujo, and K. Johansson, "Wireless networked control system co-design," in *Proc. Intl. Conf. on Networking, Sensing and Control*, Delft, The Netherlands, 2011, pp. 486–491.