Kalman filtering with intermittent heavy tailed observations

Sabina Zejnilović

Abstract

In large wireless sensor networks, data can experience loss and significant delay which from the aspect of control purposes, has the same effect as the loss. This paper reviews results achieved for Kalman filtering with intermittent observations. It has been shown that by modeling the arrival of observations as random Bernoulli process, there are conditions under which the state error covariance remains bounded. We show that if we model the arrivals with heavy tailed distribution, often used to characterize network traffic, theoretical bounds for error covariance and critical value of observation arrival still apply.

I Introduction

Applications of wireless sensor networks are numerous and they range from environmental and industrial monitoring to military surveillance and object tracking. For many of these applications, observed data is used to estimate the state of a controlled system and this estimate is later used for control. Thus, it is critical that data arrives in time. However, in large multi-hop networks, loss and delay can occur along unreliable wireless channels. Significant delay under these conditions is treated the same as the data loss. Estimation with some of the observation missing represents a topic of interest dating from the seventies. [1] [2]

For a Kalman filter estimation with intermittent observations, in a discrete system, observations can be modeled as Bernoulli random process with parameter $0 < \lambda < 1$. If the probability of arrival of an observation at time t is $\lambda > \lambda_c$ then the expectation of the estimation error covariance is bounded. The value of λ_c depends on the eigenvalues of the state matrix A and on the structure of output matrix C. If $\lambda \leq \lambda_c$ then the expectation of error covariance is not finite. This critical value of the probability of arrival of an observation λ_c cannot be directly calculated, but there exist upper and lower bounds for it. For some systems, the bounds coincide and exact value of λ_c can be evaluated. For $\lambda > \lambda_c$, upper and lower bound for the mean error covariance can be derived. These results represent the work published in [1].

Heavy-tailed distribution have been widely used to model traffic in wired and wireless networks. [4] [3] [5] Burstiness of data loss can be reflected by modeling the data inter-arrival as a random Pareto process. By evaluating λ_a, the average probability of arrival of observation that obeys heavy tailed distribution, we can apply results for convergence of error covariance for the case of $\lambda_a > \lambda_c$.

The rest of the paper is organized as follows. In section II, the formulation of problem of Kalman filtering with intermittent observations is presented. Section III presents the conditions under which the error remains bounded. Overview of heavy tailed distributions is given in section IV. In section V simulation results are presented, while in section VI conclusions are stated and future work is discussed.

II Problem statement

Discrete-time linear system with Gaussian noise assumptions has the following representation:

$$\begin{aligned} x_{t+1} &= Ax_t + w_t \\ y_t &= Cx_t + v_t \end{aligned} \tag{1}$$

where $x_t \in \Re^n$ is the state vector, $y_t \in \Re^m$ the output vector, $w_t \in \Re^p$ the plant noise and $v_t \in \Re^m$ the measurement noise. Both plant and measurement noise are independent Gaussian random vectors with zero mean and covariance matrices Q and R, respectively. With the initial state x_0 being zero mean Gaussian with covariance Σ_0, Kalman filter is an optimal estimator.

Kalman filter predict and update equations can be modified for the case of intermittent observations as follows. The arrival of the observation at time t is modeled as a binary random variable γ_t with probability distribution $p_{\gamma_t}(1) = \lambda_t$. The measurement noise v_t is defined:

$$p(v_t \mid \gamma_t) = \begin{cases} \mathcal{N}(0, R), & \gamma_t = 1 \\ \mathcal{N}(0, \sigma^2 I), & \gamma_t = 0 \end{cases}$$

The absence of observations corresponds to the case of $\sigma \to \infty$. First, the following is defined:

$$\begin{aligned} \hat{x}_{t|t} &\triangleq \mathbb{E}[x_t | \mathbf{y_t}, \gamma_\mathbf{t}] \\ P_{t|t} &\triangleq \mathbb{E}\left[(x_t - \hat{x}_t)(x_t - \hat{x}_t)' \mid \mathbf{y_t}, \gamma_\mathbf{t}\right] \\ \hat{x}_{t+1|t} &\triangleq \mathbb{E}[x_{t+1} \mid \mathbf{y_t}, \gamma_\mathbf{t}] \\ P_{t+1|t} &\triangleq \mathbb{E}\left[(x_{t+1} - \hat{x}_{t+1})(x_{t+1} - \hat{x}_{t+1})' \mid \mathbf{y_t}, \gamma_\mathbf{t}\right] \\ \hat{y}_{t+1|t} &\triangleq \mathbb{E}[y_{t+1} \mid \mathbf{y_t}, \gamma_\mathbf{t}] \end{aligned}$$

The Kalman filter equations are modified into following:

$$\begin{aligned} \hat{x}_{t+1|t} &= A\hat{x}_{t|t} \\ P_{t+1|t} &= AP_{t|t}A' + Q \qquad (2) \\ \hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + P_{t+1|t}C' \\ &\times \left(CP_{t+1|t}C' + \gamma_{t+1}R + (1-\gamma_{t+1})\sigma^2 I\right)^{-1} \\ &\times \left(y_{t+1} - C\hat{x}_{t+1|t}\right) \qquad (3) \\ Px_{t+1|t+1} &= P_{t+1|t} - P_{t+1|t}C' \\ &\times \left(CP_{t+1|t}C' + \gamma_{t+1}R + (1-\gamma_{t+1})\sigma^2 I\right)^{-1} \\ &\times CP_{t+1|t} \qquad (4) \end{aligned}$$

Now, taking the limit as $\sigma \to \infty$, the update equations (3) and (4) become:

$$\begin{aligned} \hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + \gamma_{t+1}K_{t+1}\left(y_{t+1} - C\hat{x}_{t+1|t}\right) \\ P_{t+1|t+1} &= P_{t+1|t} - \gamma_{t+1}K_{t+1}CP_{t+1} \\ K_{t+1} &= P_{t+1|t}C'\left(CP_{t+1|t}C' + R\right)^{-1} \qquad (5) \end{aligned}$$

The predict cycle remains the same as in standard case. Updated state estimate and covariance are now a a function of a random variable γ_{t+1} and therefore are also a random variable. Equations (5) mean that an update at time t occurs only if $\gamma_t = 1$, i.e. the observation has arrived. For the other case, when observation is missing, $\gamma_t = 0$, the updated value is equal to the predicted value. Given the observations $\{y_t\}$ and their arrival sequence $\{\gamma_t\}$, update equations (5) give the minimum state error covariance. Overview of the system is given on Figure 1.

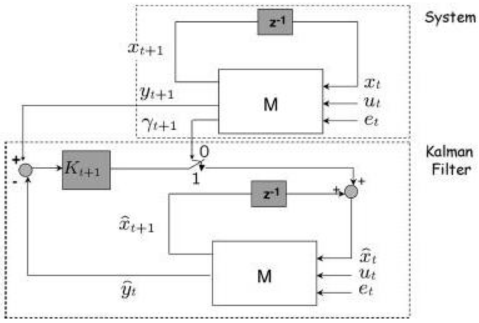


Figure 1: Kalman filter with intermittent observations

Under the assumption that the arrival process of the observation is time-independent $\lambda_t = \lambda$ for all t, there exist upper and lower bound on the expectation of the state error covariance. These bounds are a function of the arrival probability of the observation λ.

III Bounds for convergence

The Riccati equation of the state error covariance using (2) and (5) can be written in the following form:

$$P_{t+1} = AP_tA' + Q - \lambda AP_tC'\left(CP_tC' + R\right)^{-1} CP_tA' \qquad (6)$$

where $P_t = P_{t|t-1}$.

The following theorem states the existence of a critical value for the observation arrival rate below which the state error covariance may diverge for some initial condition.

Theorem 1. *If $\left(A, Q^{\frac{1}{2}}\right)$ is controllable, (A, C) is observable and A is unstable, there exists a $\lambda_c \in (0, 1]$ such that:*

$$\begin{aligned} &\lim_{t\to\infty} \mathbb{E}[P_t] = +\infty, && \text{for } 0 \le \lambda \le \lambda_c \text{ and } \exists P_0 \ge 0 \\ &\mathbb{E}[P_t] \le M_{P_0}\ \forall t && \text{for } \lambda_c < \lambda \le 1 \text{ and } \forall P_0 \ge 0 \end{aligned}$$

where $M_{P_0} > 0$ depends on the initial condition $P_0 > 0$.

The critical value cannot be directly calculated, however its upper $\overline{\lambda}$ and lower bound $\underline{\lambda}$ can be derived: $\underline{\lambda} \leq \lambda_c \leq \overline{\lambda}$.

Lower bound can easily be calculated:

$$\underline{\lambda} = 1 - \frac{1}{\alpha^2} \tag{7}$$

where $\alpha = \max_i |\sigma_i|$ and σ_i are the eigenvalues of A.

The upper bound $\overline{\lambda}$ is obtained by solving the following optimization problem:

$$\overline{\lambda} = \text{argmin}\Phi_\lambda\left(Y, Z\right) > 0, \quad 0 \leq Y \leq I$$

where function $\Phi_\lambda\left(Y, Z\right)$ is defined:

$$\Phi(Y,Z) = \begin{bmatrix} Y & \sqrt{\lambda}(YA+ZC) & \sqrt{1-\lambda}YA \\ \sqrt{\lambda}\left(YA'+C'Z'\right) & Y & 0 \\ \sqrt{1-\lambda}A'Y & 0 & Y \end{bmatrix} > 0 \tag{8}$$

This problem can be solved using bisection for λ whose upper and lower bound are known, 0 and 1, and solving convex feasibility problem with fixed λ for each iteration.

The following theorem states an estimate of the limit of the expected value $\mathbb{E}\left[P_t\right]$, when it is bounded.

Theorem 1. *If $\left(A, Q^{\frac{1}{2}}\right)$ is controllable, (A, C) is observable and $\lambda \geq \overline{\lambda}$ then the following applies:*

$$0 < S_t \leq \mathbb{E}\left[P_t\right] \leq V_t, \quad \forall \mathbb{E}\left[P_0\right] \geq 0 \tag{9}$$

with $\lim_{t\to\infty} S_t = \overline{S}$ and $\lim_{t\to\infty} V_t = \overline{V}$, where $\overline{S}$ and $\overline{V}$ are the solutions of algebraic equations:

$$\begin{aligned} \overline{S} &= \left(1 - \lambda\right) A\overline{S}A' \\ \overline{V} &= g_\lambda\left(\overline{V}\right) \\ &= A\overline{V}A' + Q - \lambda A\overline{V}C'\left(C\overline{V}C' + R\right)^{-1} C\overline{V}A' \end{aligned}$$

The solution to $\overline{V} = g_\lambda\left(\overline{V}\right)$ is also given by the following optimization problem:

$$\begin{gathered} argmax_V \; Trace\left(V\right) \\ subject\, to \\ \begin{bmatrix} AVA' - V & \sqrt{\lambda}AVC' \\ \sqrt{\lambda}CVA' & CVC' + R \end{bmatrix} \geq 0, \\ V \geq 0. \end{gathered} \tag{10}$$

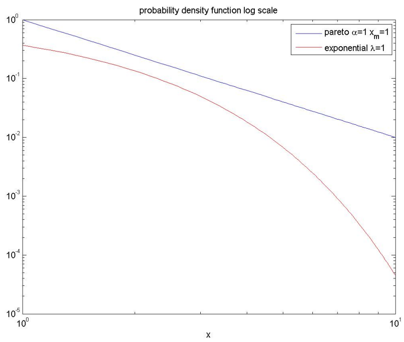


Figure 2: Pareto and exponential probability density functions plotted on a log scale

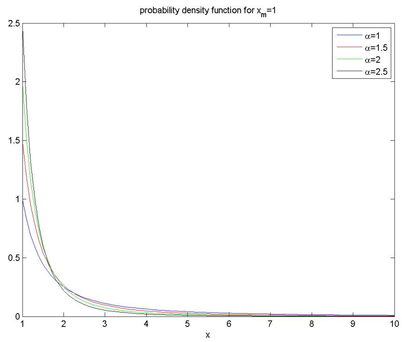


Figure 3: Pareto probability density function for various α

IV Modeling with Pareto distribution

Modeling arrival of observed data as Bernoulli variable does not take into account the burstiness of network traffic loss. Whether or not the data is lost depends only on average probability of arrival λ and not directly on whether the previous observation was lost or not. However, studies of network traffic have shown that if the previous packet was lost, then the probability that the next packet is lost is higher, meaning that data loss happens in bursts. This behavior can be modeled with heavy tailed distribution, where inter-error gaps follow Pareto distribution. Although causes for loss are different, bit errors on telephone channels and losses in Ethernet and wireless networks all show heavy tail properties.[5] [6] [7]

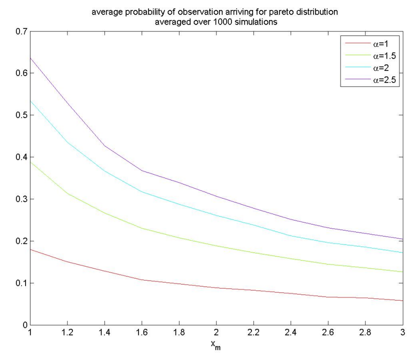


Figure 4: Average probability of arrival for Pareto distribution

Heavy tailed distributions are considered those that decay more slowly than exponential distribution, as illustrated for Pareto distribution on Figure 2. This makes larger values more probable than in exponential distribution, i.e. gives a heavier tail. Pareto distribution is characterized by shape parameter α and location parameter x_m. It's probability density function is given by:

$$f_X\left(x\right) = \left\{ \begin{array}{cc} \alpha\frac{x_m{}^{\alpha}}{x^{\alpha+1}} & for\, x > x_m \\ 0 & for\, x < x_m \end{array} \right.$$

Figure 3 shows probability density function of Pareto distribution for various α. For $\alpha \leq 1$ the expected value does not exist and for $\alpha \leq 2$ the variance does not exist.

Pareto distribution can be used to model periods between arrivals of observations. To apply previously reviewed results on convergence of error covariance, we need to use the average probability of arrival of observation. This average probability is not directly a parameter of Pareto distribution, but it can be obtained during simulations. Figure 4 shows the obtained graphs of average probability of arrival for various values of Pareto parameters α and x_m.

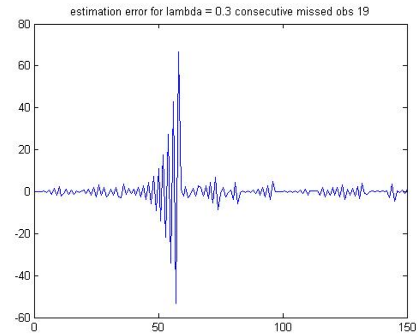
From previous results, it can be stated that the expected value $\mathbb{E}\left[P_t\right]$ should be bounded for all initial conditions P_0, if the average arrival probability of the observations update is greater then critical value λ_c, and it may diverge for some initial condition if it is lower, and this can be verified with Monte-Carlo simulations.

V Simulation results

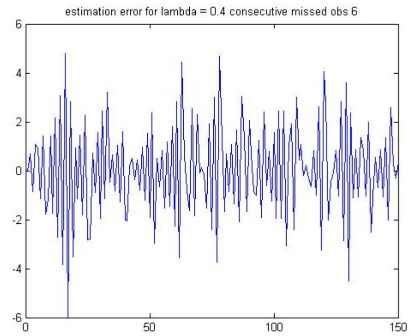
Without loss of generality, we will use a scalar version of unstable system with the following parameters: $A = -1.25$, $C = 1$, and v_t and w_t having zero mean and variance $R = 2.5$ and $Q = 1$, respectively. Using (7) and (8), lower and upper bound for critical value can be calculated. In this case, the bounds coincide and we can obtain a tight estimate of $\lambda_c = 0.36$. Figure 5(a) shows estimation error for the case when λ is below and Figure 5(b) when it is above λ_c. The error has dropped 14 times for a small increase in λ. Similarly, there is significant decrease in the value of error variance, once λ increases a little over λ_c, as seen on Figure 6.

Now, we use Pareto distribution (rounded to nearest integer) to model periods between arrivals of observations for the same system. For each simulation run, we calculate the average arrival probability λ. Figure 7(a) shows estimation error for the case when this average probability is below critical value of $\lambda_c = 0.36$, for a single simulation, with Pareto parameters ($x_m = 1$ and $\alpha = 1.35$). Figure 7(b) shows the results when multiple simulations where run with an average value of λ greater than λ_c. Absolute error was averaged for 100 simulations for the same system and the same Pareto parameters ($x_m = 1$ and $\alpha = 1.6$). Our results resemble the case of modeling observations with Bernoulli random variable where for $\lambda < \lambda_c$ estimation error can be higher couple of orders of magnitude than in the case $\lambda > \lambda_c$, even if the change in λ is not big.

Next, we compare the error covariance for the two cases of λ. Figure 8(a) shows the value of variance when average probability is below critical value, for a single simulation. Next, as for estimation error, we ran multiple simulations and averaged the variance. Then we compared the mean value in time of thus obtained variance. After solving (10) for λ equal to the average probability of arrival, we obtained an upper bound for the error variance. As it can be seen from Figure 8(b), the expected value of the variance remains below the upper bound. This bound does not apply to the value that variance can take in time, only to its expected value, as the figure shows these values can be much higher then the upper bound.

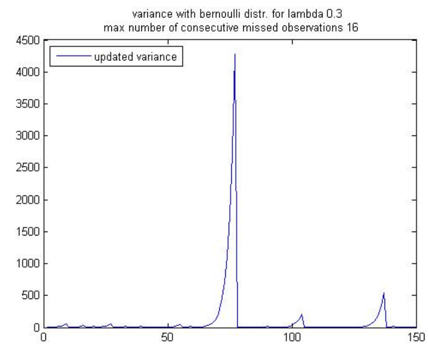


(a) λ below λ_c

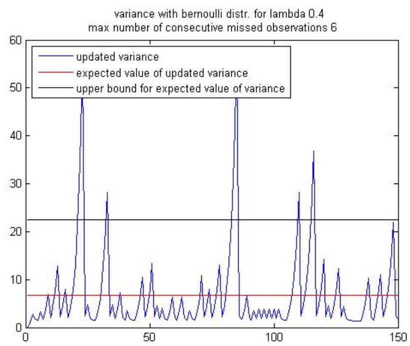


(b) λ above λ_c

Figure 5: Estimation error-observations modeled as Bernoulli random variable

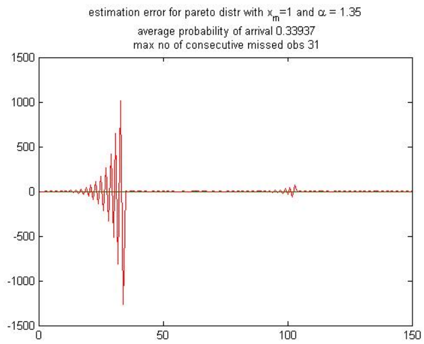


(a) λ below λ_c

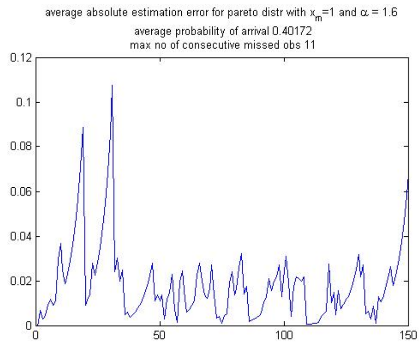


(b) λ above λ_c

Figure 6: State error variance-observations modeled as Bernoulli random variable



(a) λ below λ_c

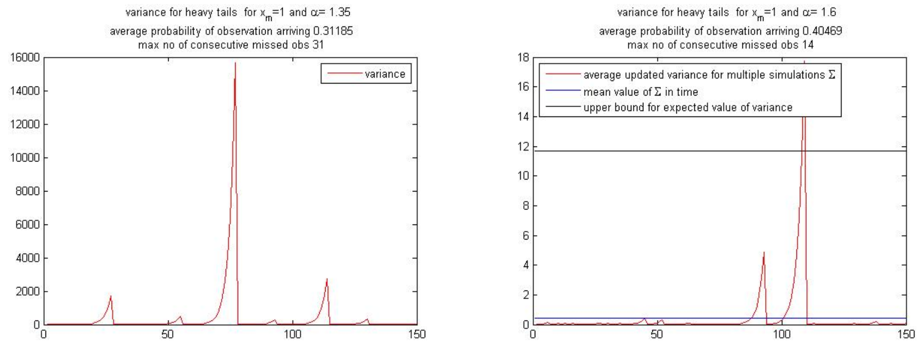


(b) λ above λ_c

Figure 7: Estimation error-observations modeled as Pareto random variable

VI Conclusions

In this paper, we have reviewed the results on Kalman filtering when not all observation data is available. Observations are modeled as Bernoulli random variables. There exists a critical value of



(a) λ below λ_c

(b) λ above λ_c

Figure 8: State error variance-observations modeled as Pareto random variable

arrival probability of observations that depends on the system dynamics, above which state covariance remains bounded for all initial conditions.

When these results are applied for observations that are modeled with Pareto distribution, which is heavy tailed, simulations lead to the same conclusions. If average arrival probability of Pareto modeled observations falls above the critical value of arrival for the unstable system, error covariance stays finite.

This experimental methodology is the first step for deriving specific bounds for the critical value of average arrival probability and error covariance for heavy-tailed characterized intermittent observations.

References

[1] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M.I. Jordan and S. S. Sastry, "Kalman Filtering With Intermittent Observations", *IEEE Transactions on automatic control*, vol. 49, no. 9, Sep 2004.

[2] M. Micheli and M. I. Jordan, "Random sampling of a continuous-time stochastic dynamical system", *15th Int. Symp. Mathematical Theory of Networks and Systems (MTNS), South Bend, IN,* Aug 2002.

[3] J. Zhao and R. Govindan, "Understanding Packet Delivery Performance In Dense Wireless Sensor Networks", *SenSys'03, Proceedings of the First International Conference on Embedded Networked Sensor Systems* Nov 2003.

[4] S.A. Mushtaq and A.A. Rizvi, "Statistical analysis and mathematical modeling of network (segment) traffic", *Proceedings of the IEEE Symposium on Emerging Technologies 2005* Sep 2005.

[5] I. Lee and A. Fapojuwo, "Estimating Heavy-tails in Long-range Dependent Wireless Traffic" *Vehicular Technology Conference* Spring 2005.

[6] J. M. Berger and B. Mandelbrot, " A New Model for Error Clustering in Telephone Circuits, IBM J RD", Jul 1963

[7] M. S. Borella and D. Swider, "Internet packet loss: Measurement and implications for end-to-end QoS", *1998 ICPP workshops on architectural and OS support for multimedia applications/exible communication systems/wireless networks and mobile computing*, Aug 1998.

[8] J.P. Nolan, "Stable Distributions – Models for Heavy Tailed Data" Chap 1, May, 2009