Submitted, Automatica, Aug 2008
http://www.cds.caltech.edu/~murray/preprints/2008l_sxm08.html

# Kalman Filtering Over A Packet-delaying Network: A Probabilistic Approach ⋆

Ling Shi [a], Lihua Xie [b], Richard M. Murray [c]

[a] *Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91106*

[b] *School of Electrical and Electrical Engineering, Nanyang Technological University, Singapore*

[c] *Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91106*

**Abstract**

In this paper, we consider Kalman filtering over a packet-delaying network. Given the probability distribution of the delay, we can completely characterize the filter performance via a probabilistic approach. We assume the estimator maintains a buffer of length $D$ so that at each time $k$, the estimator is able to retrieve all available data packets up to time $k-D+1$. Both the cases of sensor with and without necessary computation capability for filter updates are considered. When the sensor has no computation capability, for a given $D$, we give lower and upper bounds on the probability for which the estimation error covariance is within a prescribed bound. When the sensor has computation capability, we show that the previously derived lower and upper bounds are equal to each other. An approach for determining the minimum buffer length for a required performance in probability is given and an evaluation on the number of expected filter updates is provided. Examples are provided to demonstrate the theory developed in the paper.

*Key words:* Kalman filters; Networked control systems; Packet-delaying networks; Estimation theory; Probabilistic performance;

## 1 Introduction

The Kalman filter has played a central role in systems theory and has found wide applications in many fields such as control, signal processing, and communications. In the standard Kalman filter, it is assumed that sensor data are transmitted along perfect communication channels and are available to the estimator either instantaneously or with some fixed delays and no interaction between communication and control is considered. This abstraction has been adopted until recently when networks, especially wireless networks, are used in sensing and control systems for transmitting data from sensor to controller and/or from controller to actuator. While having many advantages such as low cost and flexibility, networks also induce many new issues due to their limited capabilities and uncertainties such as limited bandwidth, packet losses, and latency. On the other hand, in wireless sensor networks, sensor nodes also have limited computation capability in addition to their limitations in communications. These constraints undoubtedly affect system performance or even stability and cannot be neglected when designing estimation and control algorithms, which has inspired a lot of research in control with communication constraints; see the survey [1] and the references therein.

In recent years, networked control problems have gained much interest. In particular, the state estimation problem over a network has been widely studied. The problem of state estimation and stabilization of a linear time invariant(LTI) system over a digital communication channel which has a finite bandwidth capacity was introduced by Wong and Brockett [2, 3] and further pursued by others (e.g., in [4–7]). Sinopoli et al. [8] discussed how packet loss can affect state estimation. They showed there exists a certain threshold of the packet loss rate above which the state estimation error diverges in the expected sense, i.e., the expected value of the error covariance matrix becomes unbounded as time goes to infinity. They also provided lower and upper bounds of the

⋆ The work by L. Shi and R. M. Murray is supported in part by AFOSR grant FA9550-04-1-0169. The work by L. Xie is supported by A*STAR SERC grant 052 101 0037.

*Email addresses:* shiling@cds.caltech.edu (Ling Shi), elhxie@ntu.edu.sg (Lihua Xie), murray@cds.caltech.edu (Richard M. Murray).

threshold value. Following the spirit of [8], Liu and Goldsmith [9] extended the idea to the case where there are multiple sensors and the packets arriving from different sensors are dropped independently. They provided similar bounds on the packet loss rate for a stable estimate, again in the expected sense. [10, 11] characterize packet losses as a Markov chain and give some sufficient and necessary stability conditions under the notion of peak covariance stability. The drawback of using mean covariance matrix as a stability measure is that it may conceal the fact that events with arbitrarily low probability may make the mean value diverge. Different from [8, 10, 11], Shi et al. [12] investigates the stability of the Kalman filter via a probabilistic approach.

The problem of the Kalman filtering for systems with delayed measurements is not new and has been studied even before the emergence of networked control [13, 14]. It has been well known that discrete-time systems with constant or known time-varying bounded measurement delays may be handled by state augmentation in conjunction with the standard Kalman filtering or by the reorganized innovation approach in [15–17]. Although sensor data are usually time-stamped and thus transmission delays are known to the filter, the delays in networked systems are random in nature. Thus, the state augmentation and the reorganized innovation approaches are generally not applicable.

For the problem of randomly delayed measurements, Ray et al. [13] present a modification of the conventional minimum variance state estimator to accommodate the effects of the random arrival of measurements whereas a suboptimal filter in the least mean square sense is given in [14]. In [18], a recursive minimum variance state estimator is presented for linear discrete-time partially observed systems where the observations are transmitted by communication channels with randomly independent delays. Using covariance information, recursive least-squares linear estimators for signals with random delays are studied in [19]. Furthermore, the filtering problems with random delays and missing measurements have been investigated in [20–22] via the linear matrix inequality and the Riccati equation approaches, respectively. Note that most of the aforementioned work is concerned with the optimal or suboptimal average design where the mean filtering error covariance is taken with respect to a random i.i.d. variable that characterizes the random delay in addition to the process and measurement noises and the initial state. Thus, the derived filter is in fact suboptimal when the delay is known on-line. There has been no systematic analysis on the performance of the Kalman filter which offers the optimal filtering performance for systems with random measurement delay available on-line.

The goal of the present work is to study the performance of Kalman filter under random measurement delay. We assume that the probability distribution of the delay is given and aim to give a complete characterization of filter performance by a probabilistic approach. Due to the limited computation capability of the filtering center and also in consideration of the fact that a late arriving measurement related to the system state in the far past may not contribute much to the improvement of the accuracy of the current estimate, it is practically important to determine a proper buffer length for measurement data within which a measurement will be used to update the current state and beyond which the data will be discarded. The buffer provides a tradeoff between performance and computational load. In the paper, for a given buffer length, we shall give lower and upper bounds for the probability at which the filtering error covariance is within a prescribed bound, i.e., $\mathbf{Pr}[P_k \le M]$ for some given $M$. The upper and lower bounds can be easily evaluated by the probability distribution of the delay and the system dynamics. An approach for determining the minimum buffer length for a required performance in probability is given and an evaluation on the number of expected filter updates is provided. Both the cases of sensor with and without necessary computation capability for filter updates are considered. To the best of our knowledge, the present paper is the first in dealing with the Kalman filtering performance analysis in networked systems with random measurement delay. Our results will have both theoretical and practical importance in networked sensing and control.

The rest of the paper is organized as follows. In Section 2, the mathematical models of the problem are given. In Section 3, we consider the case when measurement data is sent via the delaying network, and we provide lower and upper bounds for $\mathbf{Pr}[P_k \le M]$. In Section 4, we consider the case when sensor estimate is sent via the delaying network, and we show that the previously derived lower and upper bounds equal to each other and hence give an exact form of $\mathbf{Pr}[P_k \le M]$. Examples are provided in Section 5 to demonstrate the theory developed in the paper. Some concluding remarks are given in the end.

## 2 Problem Setup

### *2.1 System Model*

We consider the problem of state estimation over a packet-delaying network as seen from Fig. 1. The process dynamics and sensor measurement equation are given as follows:

$$x_k = Ax_{k-1} + w_{k-1}, \tag{1}$$

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

In the above equations, $x_k \in \mathbb{R}^n$ is the state vector, $y_k \in \mathbb{R}^m$ is the observation vector, $w_{k-1} \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are zero mean white Gaussian random vectors

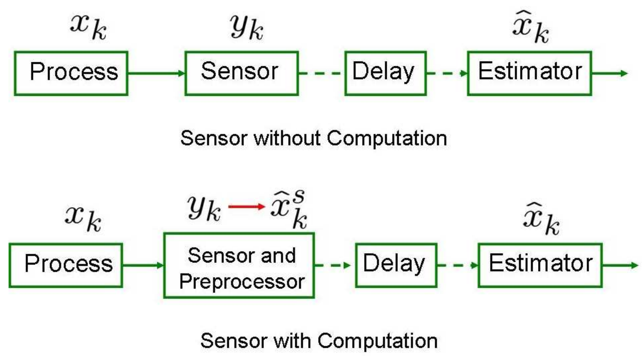


Fig. 1. System Block Diagram

with $\mathbb{E}[w_k w_j{}'] = \delta_{kj} Q, Q \geq 0$, $\mathbb{E}[v_k v_j{}'] = \delta_{kj} R, R > 0$, $\mathbb{E}[w_k v_j{}'] = 0 \; \forall j, k$, where $\delta_{kj} = 0$ if $k \neq j$ and $\delta_{kj} = 1$ otherwise. We assume that the pair $(A, C)$ is observable, $(A, \sqrt{Q})$ is controllable.

Depending on its computational capability, the sensor can either send $y_k$ or preprocess $y_k$ and send $\hat{x}_k^s$ to the remote estimator, where $\hat{x}_k^s$ is defined at the sensor as

$$\hat{x}_k^s \triangleq \mathbb{E}[x_k | y_1, \cdots, y_k].$$

The two cases correspond to the two scenarios in Fig. 1, i.e., sensor without/with computation capability.

### 2.2 Network Delay Model

After taking a measurement at time $k$, the sensor sends $y_k$ (or $\hat{x}_k^s$) to a remote estimator for generating the state estimate. We assume that the measurement data packets from the sensor are to be sent across a packet-delaying network, with negligible quantization effects, to the estimator. Each $y_k$ (or $\hat{x}_k^s$) is delayed by $d_k$ times, where $d_k$ is a random variable described by a probability mass function $f$, i.e.,

$$f(j) = \mathbf{Pr}[d_k = j], j = 0, 1, \cdots \tag{3}$$

For simplicity, we assume $d_{k_1}$ and $d_{k_2}$ are independent if $k_1 \neq k_2$. Similar to [23], we can use a markov chain to model consecutive data packet delays, and the results extend straightforward to that case. Notice that the i.i.d packet drop with drop rate $1 - \gamma$ considered in the literature can be treated as a special case here, i.e.,

$$f(0) = \gamma, f(\infty) = 1 - \gamma, f(j) = 0, 1 \leq j < \infty.$$

Thus the theory developed in the paper includes the packet drop analysis as well.

### 2.3 Problems of Interest

Define the following state estimate and other quantities at the remote estimator

$$\begin{aligned} \hat{x}_k^- &\triangleq \mathbb{E}[x_k | \text{all data packets up to } k-1], \\ \hat{x}_k &\triangleq \mathbb{E}[x_k | \text{all data packets up to } k], \\ P_k^- &\triangleq \mathbb{E}[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)'], \\ P_k &\triangleq \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)']. \end{aligned}$$

Assume the estimator discards any data $y_k$ (or $\hat{x}_k^s$) that are delayed by $D$ times or more. Given the system and the network delay models in Eqn (1)-(3), we are interested in the following problems.

(1) How should $\hat{x}_k$ be computed?
(2) What is the relationship between $P_k$ and $D$?
(3) For a given $M \geq 0$ and $\epsilon \in [0, 1]$, what is the minimum $D$ such that

$$\mathbf{Pr}[P_k \leq M | D] \geq 1 - \epsilon.$$

In the rest of the paper, we provide solutions to the above three problems for each of the two scenarios in Fig. 1.

The following terms that are frequently used in subsequent sections are defined below. It is assumed that $(A, C, Q, R)$ are the same as they appear in Section 2; $\lambda_i(A)$ is the $i$th eigenvalue of the matrix $A$; $X \in \mathbb{S}_+^n$ where $\mathbb{S}_+^n$ is the set of $n$ by $n$ positive semi-definite matrices; $h, g : \mathbb{S}_+^n \to \mathbb{S}_+^n$ are functions defined below; $Y_i$ is a random variable where the underlying sample spaces will be clear from its context.

$$\begin{aligned} \rho(A) &\triangleq \max_i |\lambda_i(A)| \\ h(X) &\triangleq AXA' + Q \\ g(X) &\triangleq h(X) - AXC'[CXC' + R]^{-1} CXA' \\ \tilde{g}(X) &\triangleq X - XC'[CXC' + R]^{-1} CX \\ h \circ g(X) &\triangleq h(g(X)) \\ \mathbf{Pr}[Y_1 | Y_2] &\triangleq \mathbf{Pr}[Y_1] \text{ given } Y_2 \end{aligned}$$

## 3 Sensor without Computation Capability

In this section, we consider the first scenario in Fig. 1, i.e., the sensor has no computation and sends $y_k$ to the remote estimator. We assume $C$ is full rank, and without loss of generality, we assume $C^{-1}$ exists. The general $C$ case will be considered in Section 4.1.

### 3.1 Modified Kalman Filtering

Let $\gamma_t^k$ be the indicator functor for $y_t$ at time $k, t \leq k$, which is defined as follows.

$$\gamma_t^k = \begin{cases} 1, y_t \text{ received at time } k, \\ 0, \text{otherwise.} \end{cases}$$

Further define $\gamma_{k-i} \triangleq \sum_{j=0}^{i} \gamma_{k-i}^{k-j}$, i.e., $\gamma_{k-i}$ indicates whether $y_{k-i}$ is received by the estimator at or before $k$.

Assume $\hat{x}_{k-1}$ is optimal. Depending on whether $y_k$ is received or not, i.e., $\gamma_k^k = 1$ or 0, $(\hat{x}_k, P_k)$ is known to be computed by a Modified Kalman Filter (**MKF**) [8]. We write $(\hat{x}_k, P_k)$ in compact form as follows.

$$(\hat{x}_k, P_k) = \mathbf{MKF}(\hat{x}_{k-1}, P_{k-1}, \gamma_k^k y_k)$$

which represents the follow set of equations:

$$\begin{cases} \hat{x}_k^- &= A\hat{x}_{k-1}, \\ P_k^- &= AP_{k-1}A' + Q, \\ K_k &= P_k^- C'[CP_k^- C' + R]^{-1}, \\ \hat{x}_k &= A\hat{x}_{k-1} + \gamma_k^k K_k (y_k - CA\hat{x}_{k-1}), \\ P_k &= (I - \gamma_k^k K_k C)P_k^-. \end{cases}$$

Assume $\gamma_k^k = 1$ for all $k$, then **MKF** reduces to the standard Kalman filter. In this case, $P_k^-$ and $P_k$ can be shown to satisfy

$$P_k^- = g(P_{k-1}^-), \quad P_k = \tilde{g} \circ h(P_{k-1}).$$

Let $P^*$ be the unique positive semi-definite solution [1] to $g(X) = X$, i.e., $P^* = g(P^*)$. Define $\overline{P}$ as $\overline{P} \triangleq \tilde{g}(P^*)$. Then we have

$$\tilde{g} \circ h(\overline{P}) = \tilde{g} \circ h \circ \tilde{g}(P^*) = \tilde{g} \circ g(P^*) = \tilde{g}(P^*) = \overline{P},$$

where we use the fact that $h \circ \tilde{g} = g$. In other words,

$$P^* = \lim_{k \to \infty} P_k^-, \quad \overline{P} = \lim_{k \to \infty} P_k.$$

### 3.2 Optimal Estimation with Delayed Measurements

As $y_{k-i}$ may arrive at time $k$ due to the delays introduced by the network, we can improve the estimation quality by recalculating $\hat{x}_{k-i}$ utilizing the new available measurement $y_{k-i}$. Once $\hat{x}_{k-i}$ is updated, we can update $\hat{x}_{k-i+1}$ in a similar fashion. This is the basic idea contained in the flow diagram in Fig. 2. Theorem 3.1 summarizes the main estimation result.

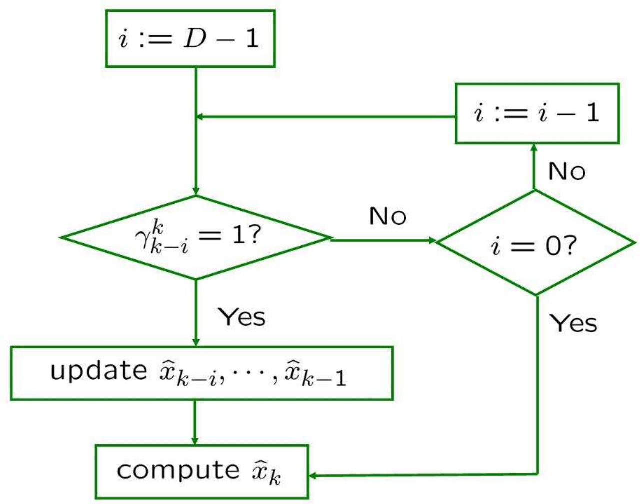


Fig. 2. Optimal Estimation: Sensor without Computation Capability

**Theorem 3.1** *Let $y_{k-i}, i \in [0, D-1]$ be the oldest measurement received by the estimator at time $k$. Then $\hat{x}_k$ is computed by $i+1$ Modified Kalman Filters as*

$$\begin{aligned} (\hat{x}_{k-i}, P_{k-i}) &= \boldsymbol{MKF}(\hat{x}_{k-i-1}, P_{k-i-1}, y_{k-i}) \\ (\hat{x}_{k-i+1}, P_{k-i+1}) &= \boldsymbol{MKF}(\hat{x}_{k-i}, P_{k-i}, \gamma_{k-i+1} y_{k-i+1}) \\ &\vdots \\ (\hat{x}_{k-1}, P_{k-1}) &= \boldsymbol{MKF}(\hat{x}_{k-2}, P_{k-2}, \gamma_{k-1} y_{k-1}) \\ (\hat{x}_k, P_k) &= \boldsymbol{MKF}(\hat{x}_{k-1}, P_{k-1}, \gamma_k y_k). \end{aligned}$$

*Furthermore, $\hat{D}$, the average number of **MKF** used at each time $k$ is given by*

$$\hat{D} = \prod_{i=1}^{D-1} \big(1 - f(i)\big) + \sum_{j=2}^{D} \prod_{i=j}^{D-1} \big(1 - f(i)\big) f(j-1) j, \quad (4)$$

*where*

$$\prod_{i=D}^{D-1} \big(1 - f(i)\big) \triangleq 1.$$

**Proof:** We know that the estimate $\hat{x}_k$ is generated from the estimate of $\hat{x}_{k-1}$ together with $\gamma_k y_k$ at time $k$ through a Modified Kalman Filter. Similarly, the estimate $\hat{x}_{k-1}$ is generated from the estimate of $\hat{x}_{k-2}$ together with $\gamma_{k-1} y_{k-1}$ at time $k$ through a Modified Kalman Filter, etc. This recursion for $i+1$ steps corresponds to the $i+1$ Modified Kalman Filters stated in

[1] Since $(A, C)$ is assumed to be observable and $(A, \sqrt{Q})$ controllable, from standard Kalman filtering analysis, $P^*$ exists.

the theorem. Let $\hat{D}_k$ be the number of **MKF** used at each time $k$. Notice that $1 \le \hat{D}_k \le D$. Thus

$$\hat{D} = \sum_{j=1}^{D} j \mathbf{Pr}[\hat{D}_k = j].$$

Consider $\mathbf{Pr}[\hat{D}_k = 1]$. Since $\hat{D}_k = 1$ iff $\gamma_{k-i}^k = 0$ for all $1 \le i \le D-1$, we have

$$\mathbf{Pr}[\hat{D}_k = 1] = \mathbf{Pr}[\gamma_{k-i}^k = 0, 1 \le i \le D-1].$$

As $\mathbf{Pr}[\gamma_{k-i}^k = 0] = 1 - f(i)$, we obtain

$$\mathbf{Pr}[\hat{D}_k = 1] = \prod_{i=1}^{D-1} \big(1 - f(i)\big).$$

Similarly, for $2 \le j \le D-1$, we have

$$\mathbf{Pr}[\hat{D}_k = j] = \prod_{i=j}^{D-1} \big(1 - f(i)\big) f(j-1)$$

and when $j = D$,

$$\mathbf{Pr}[\hat{D}_k = j] = f(D-1).$$

Therefore we obtain $\hat{D}$ in Eqn (4).

**Remark 3.2** *Notice that in the first **MKF**, $\gamma_{k-i}^k = 1$ and hence $\gamma_{k-i} = 1$. As a result, we simply write $\gamma_{k-i} y_{k-i} = y_{k-i}$.*

### *3.3 Lower and Upper Bounds of* $\mathbf{Pr}[P_k \le M | D]$

Since $d_k$ is random and described by the probability mass function $f$, $\gamma_{k-i}^k$ $(i = 0, \cdots, D-1)$ is also random. As a consequence, $P_k$ computed as in Theorem 3.1 is a random variable. Define $\hat{\gamma}_i(D)$ as

$$\hat{\gamma}_i(D) \triangleq \begin{cases} \sum_{j=0}^{i} f(j), \text{if } 0 \le i < D, \\ \sum_{j=0}^{D-1} f(j), \text{if } i \ge D. \end{cases}$$

Recall that $\gamma_{k-i}$ indicates whether $y_{k-i}$ is received by the estimator at or before $k$, so it is easy to verify that

$$\mathbf{Pr}[\gamma_{k-i} = 1 | D] = \hat{\gamma}_i(D). \tag{5}$$

Define $\overline{M} \triangleq C^{-1} R C^{-1'}$. Then we have the following result that shows the relationship between $P_k$ and $\overline{M}$.

**Lemma 3.3** *For any $k \ge 1$, if $\gamma_k = 1$, then $P_k \le \overline{M}$.*

**Proof:** As $\gamma_k = 1$, we have $P_k = \tilde{g} \circ h(P_{k-1}) \le \overline{M}$, where the inequality is from Lemma A.2 in Appendix A.

**Remark 3.4** *We can also interpret Lemma 3.3 as follows. One way to obtain an estimate $\tilde{x}_k$ when $\gamma_k = 1$ is simply by inverting the measurement, i.e., $\tilde{x}_k = C^{-1} y_k$. Therefore*

$$\tilde{e}_k = C^{-1} v_k \text{ and } \tilde{P}_k = \mathbb{E}[\tilde{e}_k \tilde{e}_k^{'}] = C^{-1} R C^{-1'} = \overline{M}.$$

*Since Kalman filter is optimal among the set of all linear filters, we must have $P_k \le \tilde{P}_k = \overline{M}$.*

Recall that $\overline{P}$ is defined in Section 3.1 as the steady state error covariance for the Kalman filter. For $M \ge \overline{M}$, let us define $k_1(M)$ and $k_2(M)$ as follows:

$$k_1(M) \triangleq \min\{t \ge 1 : h^t(\overline{M}) \nleq M\}, \tag{6}$$
$$k_2(M) \triangleq \min\{t \ge 1 : h^t(\overline{P}) \nleq M\}. \tag{7}$$

We sometimes write $k_i(M)$ as $k_i, i = 1, 2$ for simplicity for the rest of the paper. The following lemma shows the relationship between $\overline{P}$ and $\overline{M}$ as well as $k_1$ and $k_2$.

**Lemma 3.5** *(1) $\overline{P} \le \overline{M}$; (2) $k_1 \le k_2$ whenever either $k_i$ is finite, $i = 1, 2$.*

**Proof:** (1) $\overline{P} = \tilde{g}(P^*) \le \overline{M}$ where the inequality is from Lemma A.2 in Appendix A. (2) Without loss of generality, we assume $k_2$ is finite. If $k_1$ is finite, and $k_1 > k_2$, then according to their definitions, we must have

$$M \ge h^{k_1 - 1}(\overline{M}) \ge h^{k_1 - 1}(\bar{P}) \ge h^{k_2}(\bar{P}),$$

which violates the definition of $k_2$. Notice that we use the property that $h$ is nondecreasing as well as $h(\bar{P}) \ge \bar{P}$ from Lemma A.1 and A.3 in Section A in the Appendix. Similarly we can show that $k_1$ cannot be infinite. Therefore we must have $k_1 \le k_2$.

**Lemma 3.6** *Assume $P_0 \ge \overline{P}$. Then $P_k \ge \overline{P}$ for all $k \ge 0$.*

**Proof:** Since **MKF** is used at each time $k$,

$$P_k = \hat{f}_k^k \circ \hat{f}_{k-1}^k \cdots \hat{f}_1^k(P_0) \ge \overline{P},$$

where $\hat{f}_{k-i}^k = h$ or $\hat{f}_{k-i}^k = \tilde{g} \circ h$ depending on the packet arrival sequence [2]. The inequality is from Lemma A.1 in Appendix A.

[2] Notice that we use the superscript $k$ in $\hat{f}_{k-i}^k$ to emphasize that it depends on the current time $k$. For example, if $d_{k-i} = i + 1$, i.e., $\gamma_{k-i} = 0$ and $\gamma_{k-i}^{k+1} = 1$, then $\hat{f}_{k-i}^k = h$ and $\hat{f}_{k-i}^{k+1} = \tilde{g} \circ h$.

Define $N_k$ as the number of consecutive packets not received by $k$, i.e.,

$$N_k \triangleq \min\{t \geq 0 : \gamma_{k-t} = 1\}. \qquad (8)$$

Define

$$\theta(k_i, D) \triangleq \prod_{j=0}^{k_i-1} \big(1 - \hat{\gamma}_j(D)\big). \qquad (9)$$

It is easy to see that

$$\theta(k_1, D) \geq \theta(k_2, D).$$

**Lemma 3.7** *Let $k_1, k_2$ and $N_k$ be defined according to Eqn (6)-(8). Then*

$$\mathbf{Pr}[N_k \geq k_i | D] = \theta(k_i, D), i = 1, 2. \qquad (10)$$

**Proof:**

$$\begin{aligned}\mathbf{Pr}[N_k \geq k_i|D] &= \mathbf{Pr}[\gamma_{k-i} = 0, 0 \leq i \leq k_i - 1|D] \\ &= \theta(k_i, D).\end{aligned}$$

**Theorem 3.8** *Assume $\overline{P} \leq P_0 \leq \overline{M}$. For any $M \geq \overline{M}$, we have*

$$1 - \theta(k_1, D) \leq \mathbf{Pr}[P_k \leq M|D] \leq 1 - \theta(k_2, D). \qquad (11)$$

**Proof:** We divide the proof into two parts. For the rest of the proof, all probabilities are conditioned on the given $D$. (1) Let us first prove $1 - \theta(k_1, D) \leq \mathbf{Pr}[P_k \leq M|D]$, or in other words,

$$1 - \mathbf{Pr}[N_k \geq k_1|D] \leq \mathbf{Pr}[P_k \leq M|D].$$

As $\gamma_k = 1$ or 0, there are in total $2^k$ possible realizations of $\gamma_1$ to $\gamma_k$ as seen from Fig. 3. Let $\Sigma_1$ denote those packet arrival sequences of $\gamma_1$ to $\gamma_k$ such that $N_k \geq k_1$. Similarly let $\Sigma_2$ denote those packet arrival sequences such that $N_k < k_1$. Let $P_k(\sigma_i)$ be the error covariance at time $k$ when the underlying packet arrival sequence is $\sigma_i$, where $\sigma_i \in \Sigma_i, i = 1, 2$. Consider a particular $\sigma_2 \in \Sigma_2$. As $\gamma_{k-k_1+1} = 1$, from Lemma 3.3, $P_{k-k_1+1} \leq \overline{M}$. Therefore we have

$$P_k(\sigma_2) \leq h^{k_1-1}(P_{k-k_1+1}) \leq h^{k_1-1}(\overline{M}) \leq M,$$

where the first and second inequalities are from Lemma A.1 in Appendix A and the last inequality is from the definition of $k_1$. In other words,

$$\mathbf{Pr}[P_k \leq M|\sigma_2] = 1.$$

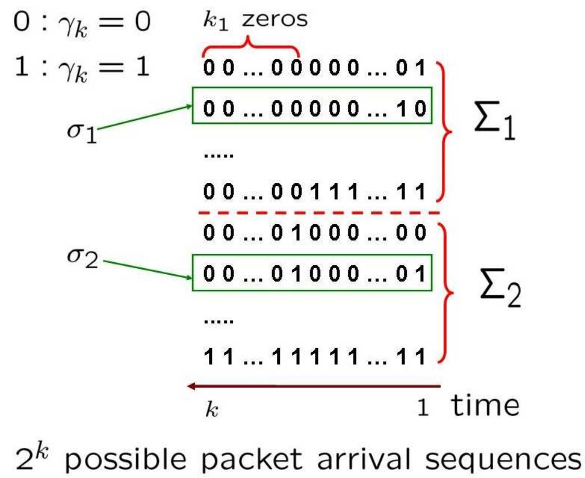


Fig. 3. $N_k \geq k_1$

Therefore we have

$$\begin{aligned}\mathbf{Pr}[P_k \leq M] &= \sum_{\sigma \in \Sigma_1 \cup \Sigma_2} \mathbf{Pr}[P_k \leq M|\sigma]\mathbf{Pr}(\sigma) \\ &= \sum_{\sigma_1 \in \Sigma_1} \mathbf{Pr}[P_k \leq M|\sigma_1]\mathbf{Pr}(\sigma_1) \\ &\quad + \sum_{\sigma_2 \in \Sigma_2} \mathbf{Pr}[P_k \leq M|\sigma_2]\mathbf{Pr}(\sigma_2) \\ &\geq \sum_{\sigma_2 \in \Sigma_2} \mathbf{Pr}[P_k \leq M|\sigma_2]\mathbf{Pr}(\sigma_2) \\ &= \sum_{\sigma_2 \in \Sigma_2} \mathbf{Pr}(\sigma_2) \\ &= \mathbf{Pr}(\Sigma_2) \\ &= 1 - \mathbf{Pr}(\Sigma_1) \\ &= 1 - \mathbf{Pr}[N_k \geq k_1],\end{aligned}$$

where the first equality is from the Total Probability Theorem, the second equality holds as $\Sigma_1$ and $\Sigma_2$ are disjoint, and the third inequality holds as the first sum is non-negative. The rest equalities are easy to see.

(2) We now prove $\mathbf{Pr}[P_k \leq M|D] \leq 1 - \theta(k_2, D)$, or in other words

$$\mathbf{Pr}[P_k \leq M|D] \leq 1 - \mathbf{Pr}[N_k \geq k_2|D].$$

Let $\Sigma_1'$ denote those packet arrival sequences of $\gamma_1$ to $\gamma_k$ such that $N_k \geq k_2$ and $\Sigma_2'$ denote those packet arrival sequences such that $N_k < k_2$ (Fig. 4). Consider $\sigma_1' \in \Sigma_1'$. Let

$$s(\sigma_1') = \min\{t \geq 1 : \gamma_{k-t} = 1|\sigma_1'\}.$$

As $\sigma_1' \in \Sigma_1'$, we must have $s \geq k_2$. Consequently,

$$P_k(\sigma_1') = h^{s(\sigma_1')}(P_{k-s(\sigma_1')}) \geq h^{s(\sigma_1')}(\overline{P}),$$

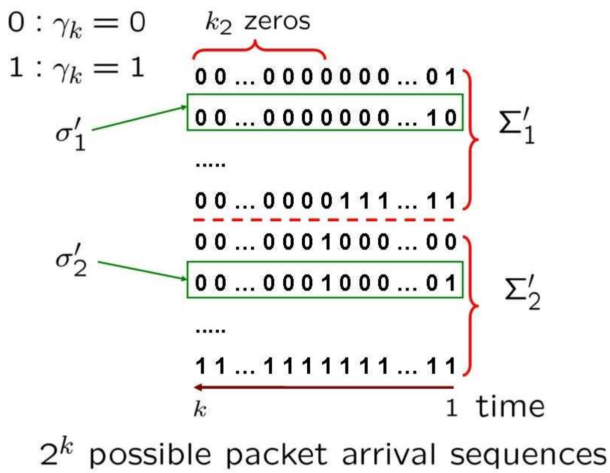


Fig. 4. $N_k \geq k_2$

where the inequality is from Lemma 3.6. Therefore we conclude that

$$P_k(\sigma_1') \nleq M.$$

Otherwise

$$h^{s(\sigma_1')}(\overline{P}) \leq P_k(\sigma_1') \leq M,$$

which violates the definition of $k_2$. In other words,

$$\mathbf{Pr}[P_k \leq M|\sigma_1'] = 0.$$

Therefore we have

$$\begin{aligned}\mathbf{Pr}[P_k \leq M] &= \sum_{\sigma \in \Sigma_1 \cup \Sigma_2} \mathbf{Pr}[P_k \leq M|\sigma]\mathbf{Pr}(\sigma) \\ &= \sum_{\sigma_1' \in \Sigma_1'} \mathbf{Pr}[P_k \leq M|\sigma_1']\mathbf{Pr}(\sigma_1') \\ &\quad + \sum_{\sigma_2' \in \Sigma_2'} \mathbf{Pr}[P_k \leq M|\sigma_2']\mathbf{Pr}(\sigma_2') \\ &= \sum_{\sigma_2' \in \Sigma_2'} \mathbf{Pr}[P_k \leq M|\sigma_2']\mathbf{Pr}(\sigma_2') \\ &\leq \sum_{\sigma_2' \in \Sigma_2'} \mathbf{Pr}(\sigma_2') \\ &= \mathbf{Pr}(\Sigma_2') \\ &= 1 - \mathbf{Pr}(\Sigma_1') \\ &= 1 - \mathbf{Pr}[N_k \geq k_2],\end{aligned}$$

where the inequality is from the fact that

$$\mathbf{Pr}[P_k \leq M|\sigma_2'] \leq 1 \text{ for any } \sigma_2' \in \Sigma_2'.$$

### *3.4 Computing the Minimum D*

Assume we require that

$$\mathbf{Pr}[P_k \leq M|D] \geq 1 - \epsilon, \tag{12}$$

then according to Eqn (11), a sufficient condition is that

$$\theta(k_1, D) \leq \epsilon. \tag{13}$$

And a necessary condition is that

$$\theta(k_2, D) \leq \epsilon. \tag{14}$$

For a given $M$, define

$$\epsilon_1(M) \triangleq \theta(k_1, k_1 - 1), \tag{15}$$

$$\epsilon_2(M) \triangleq \theta(k_2, k_2 - 1). \tag{16}$$

#### *3.4.1 Sufficient Minimum D*

Notice that $\theta(k_1, D)$ is decreasing when $1 \leq D \leq k_1 - 1$ and remains constant when $D \geq k_1$. Hence if $\epsilon < \epsilon_1(M)$, no matter how large $D$ is, there is *no guarantee* that $\mathbf{Pr}[P_k \leq M|D] \geq 1-\epsilon$. If $\epsilon \geq \epsilon_1(M)$, then the minimum $D_s$ such that *guarantees* $\mathbf{Pr}[P_k \leq M|D] \geq 1 - \epsilon$ is given by

$$D_s = \min\{D : \theta(k_1, D) \leq \epsilon, 1 \leq D \leq k_1 - 1\}. \tag{17}$$

#### *3.4.2 Necessary Minimum D*

Similarly, $\theta(k_2, D)$ is decreasing when $1 \leq D \leq k_2 - 1$ and remains constant when $D \geq k_2$. Hence if $\epsilon < \epsilon_2(M)$, no matter how large $D$ is, it is guaranteed that $\mathbf{Pr}[P_k \leq M|D] > 1 - \epsilon$. If $\epsilon \geq \epsilon_1(M)$, then the minimum $D_s$ such that it is *possible* that $\mathbf{Pr}[P_k \leq M|D] \geq 1-\epsilon$ is given by

$$D_s = \min\{D : \theta(k_2, D) \leq \epsilon, 1 \leq D \leq k_2 - 1\}. \tag{18}$$

**Example 3.9** *Consider Eqn (1) and (2) with*

$$A = 1.4, C = 1, Q = 0.2, R = 0.5.$$

*We model the packet delay as a poisson distribution with mean d, i.e., the probability density function $f(i)$ satisfies*

$$f(i) = \frac{d^i e^{-d}}{i!}, i = 0, 1, \cdots$$

*where $d = \mathbb{E}[d_k]$ denotes the mean value of the packet delay.*

*When $M = 50$, it is calculated that $k_1(M) = k_2(M) = 7$, hence $\theta(k_1, D) = \theta(k_2, D)$ and $\theta(7, 6) = 0.0313$. Thus we can find the minimum D that guarantees $\mathbf{Pr}[P_k \leq 50] \geq 1 - \epsilon$ for any $\epsilon \geq 0.0313$. For any $\epsilon < 0.0313$, no matter how large D is, $\mathbf{Pr}[P_k \leq 50|D] < 1 - \epsilon$.*

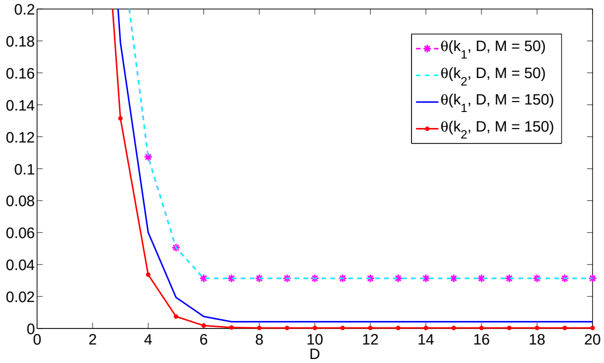


Fig. 5. $\theta(k_i, D)$ for different $M$

*When $M = 150$, it is calculated that $k_1(M) = 8$ and $k_2(M) = 9$, hence $\theta(k_1, D) > \theta(k_2, D)$. We also find that $\theta(8, 7) = 0.0042$ and $\theta(9, 8) = 0.0003$. Therefore if $\epsilon > 0.0042$, we can find minimum $D$ that guarantees $\mathbf{Pr}[P_k \leq 150] \geq 1 - \epsilon$; if $\epsilon < 0.0003$, no matter how large $D$ is, $\mathbf{Pr}[P_k \leq 150|D] > 1 - \epsilon$.*

**Remark 3.10** *We find the minimum $D$ that gives the desired filter performance, i.e., $\mathbf{Pr}[P_k \leq M|D] \geq 1-\epsilon$ for a given $M$ and $\epsilon$. We can also find the minimum $D$ when requiring $\mathbb{E}[P_k|D]$ to be stable, i.e., $\lim_{k\to\infty} \mathbb{E}[P_k|D] < \infty$, and we provide the detailed analysis in Appendix B using the theory developed in this section.*

## 4 Sensor with Computation Capability

In this section, we consider the second scenario in Fig. 1, i.e., the sensor has necessary computation capability and sends $\hat{x}_k^s$ to the remote estimator. We assume all the variables in this section, e.g., $\gamma_t^k, \gamma_k$, etc are the same as they are defined in Section 3 unless they are explicitly defined.

Consider the case when $k$ is sufficiently large so that the Kalman filter enters steady state at the sensor side, i.e., $P_k^s = \overline{P}$. It is clear that the optimal estimation at the remote estimator is as follows. If $\gamma_k = 1$, then $\hat{x}_k = \hat{x}_k^s$ and $P_k = P_k^s = \overline{P}$. If $\gamma_k = 0$ and $\gamma_{k-1} = 1$, then $\hat{x}_k = A\hat{x}_{k-1}^s$ and $P_k = h(\overline{P})$. This is repeated until we examine $\gamma_{k-D+1}$. The full optimal estimation algorithm is presented in Fig. 6.

Notice that in the first scenario (Fig. 2), i.e., sensor has no computation capability, we examine the sequence from $\gamma_{k-D+1}^k$ to $\gamma_k^k$, while in the the second scenario, (Fig. 6), we examine the sequence from $\gamma_k$ to $\gamma_{k-D+1}$.

**Theorem 4.1** *Assume $k$ is sufficiently large such that $P_k^s = \overline{P}$. For any $M \geq \overline{P}$, we have*

$$\mathbf{Pr}[P_k \leq M|D] = 1 - \theta(k_2, D). \qquad (19)$$

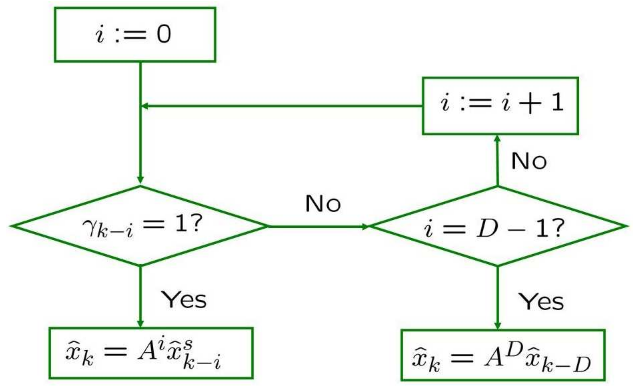


Fig. 6. Optimal Estimation: Sensor with Computation Capability

**Proof:** For the rest of the proof, all probabilities are conditioned on the given $D$. Let $\sigma_i'$ and $\Sigma_i', i = 1, 2$ be defined in the same way as in the proof of Theorem 3.8 (see Fig. 4). Clearly for any $\sigma_2' \in \Sigma_2'$,

$$P_k(\sigma_2') \leq h^{k_2-1}(\overline{P}) \leq M$$

The first inequality is from the fact that $\gamma_{k-k_2+1} = 1$ and hence $P_{k-k_2+1} = \overline{P}$. The second inequality is from the definition of $k_2$. In other words,

$$\mathbf{Pr}[P_k \leq M|\sigma_2'] = 1.$$

Similar to the proof of Theorem 3.8 , for $\sigma_1' \in \Sigma_1'$, let us define

$$s = s(\sigma_1') \triangleq \min\{t \geq 1 : \gamma_{k-t} = 1|\sigma_1'\}.$$

As $\sigma_1' \in \Sigma_1'$, $s \geq k_2$. Therefore

$$P_k(\sigma_1') = h^s(\overline{P}) \nleq M.$$

In other words, $\mathbf{Pr}[P_k \leq M|\sigma_1'] = 0$. Therefore

$$\begin{aligned}\mathbf{Pr}[P_k \leq M] &= \sum_{\sigma' \in \Sigma_1' \cup \Sigma_2'} \mathbf{Pr}[P_k \leq M|\sigma']\mathbf{Pr}(\sigma') \\ &= \sum_{\sigma_1' \in \Sigma_1'} \mathbf{Pr}[P_k \leq M|\sigma_1']\mathbf{Pr}(\sigma_1') \\ &\quad + \sum_{\sigma_2' \in \Sigma_2'} \mathbf{Pr}[P_k \leq M|\sigma_2']\mathbf{Pr}(\sigma_2') \\ &= \sum_{\sigma_2' \in \Sigma_2'} \mathbf{Pr}[P_k \leq M|\sigma_2']\mathbf{Pr}(\sigma_2') \\ &= \sum_{\sigma_2' \in \Sigma_2'} \mathbf{Pr}(\sigma_2') \\ &= \mathbf{Pr}(\Sigma_2') \\ &= 1 - \mathbf{Pr}(\Sigma_1') \\ &= 1 - \mathbf{Pr}[N_k \geq k_2].\end{aligned}$$

Computing $\mathbf{Pr}[N_k \geq k_2|D]$ follows exactly the same way as in Section 3.4. Since we have a strict equality in Eqn (19), in order that

$$\mathbf{Pr}[P_k \leq M|D] \geq 1-\epsilon$$

a necessary and sufficient condition is that

$$\mathbf{Pr}[N_k \geq k_2] \leq \epsilon. \tag{20}$$

Therefore the minimum $D^*$ that guarantees Eqn (20) to hold is given by

$$D^* = \min\{D : \theta(k_2, D) \leq \epsilon, 1 \leq D \leq k_2 - 1\}. \tag{21}$$

Notice that since $\theta(k_2, D) \geq \theta(k_2, k_2-1) = \epsilon_2(M)$, $D^*$ from the above equation exists if and only if $\epsilon \geq \epsilon_2(M)$.

### *4.1 When C Is Not Full Rank*

We use Theorem 4.1 to tackle the case when $C$ is not full rank for the first scenario, i.e., sensor without computation capability. Since $(A, C)$ is observable, there exists $r$ $(2 \leq r \leq n)$ such that

$$\begin{bmatrix} C \\ CA \\ \cdots \\ CA^{r-1} \end{bmatrix}$$

is full rank. In this section, we consider the special case when $r = 2$, and in particular, we assume $\begin{bmatrix} C \\ CA \end{bmatrix}^{-1}$ exists. The idea readily extends to the general case.

Unlike the case when $C^{-1}$ exists, and $y_k$ is sent across the network, here we assume that the previous measurement $y_{k-1}$ is sent along with $y_k$. This only requires that the sensor has a buffer that stores $y_{k-1}$. Then if $\gamma_k = 1$, both $y_k$ and $y_{k-1}$ are received. Thus we can use the following linear estimator to generate $\hat{x}_k$

$$\hat{x}_k = A \begin{bmatrix} CA \\ C \end{bmatrix}^{-1} \begin{bmatrix} y_k \\ y_{k-1} \end{bmatrix}.$$

The corresponding error covariance can be calculated as

$$P_k = AM_1A' + Q,$$

where

$$M_1 = \begin{bmatrix} CA \\ C \end{bmatrix}^{-1} \begin{bmatrix} CQC' + R & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} CA \\ C \end{bmatrix}^{-1'}.$$

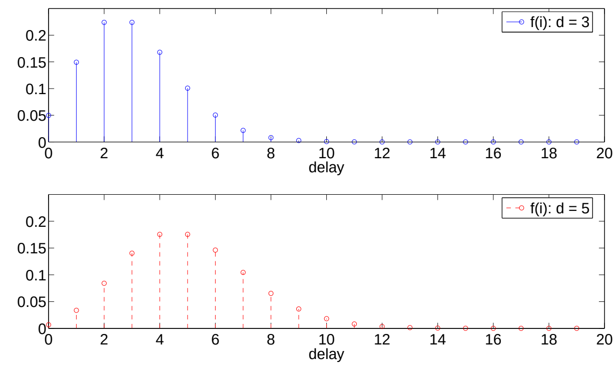


Fig. 7. Poisson distribution with $d = 3$ and $d = 5$

Therefore once the packet for time $k$ is received, i.e., $\gamma_k = 1$, we have

$$P_k = AM_1A' + Q \triangleq \tilde{P}.$$

Now if we treat $\tilde{P}$ as the steady state error covariance at the sensor side, i.e., by letting $P_k^s = \tilde{P}$, and define

$$k_v \triangleq \min\{t \geq 1 : h^t(\tilde{P}) \not\leq M\},$$

we immediately obtain

$$\mathbf{Pr}[P_k \leq M|D] = 1 - \theta(k_v, D). \tag{22}$$

**Remark 4.2** *Though we give the exact expression of* $\mathbf{Pr}[P_k \leq M|D]$ *in Eqn (22), we have to point out that* $\theta(k_v, D) \geq \theta(k_2, D)$, *as* $\tilde{P} \geq \overline{P}$ *due to the optimality of Kalman filter. Thus the case that sensor has computation capability leads to better filter performance, which is illustrated from the vector system example in the next section.*

## 5 Examples

### *5.1 Scalar System*

Consider the same parameters as in Example 3.9, i.e.,

$$A = 1.4, C = 1, Q = 0.2, R = 0.5$$

and

$$f(i) = \frac{d^i e^{-d}}{i!}, i = 0, 1, \cdots$$

Fig. 7 shows the values of $f(i)$ for $0 \leq i \leq 20$ for $d = 3$ and 5 respectively.

#### *5.1.1 Sensor without Computation Capability*

We run a Monte Carlo simulation for different parameters. Fig. 8 to 10 show the results when $D$ and $d$ take different values. From Fig. 11, we can see that both smaller

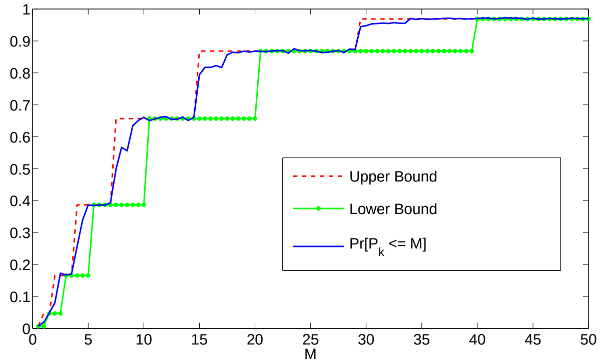


Fig. 8. $\mathbf{Pr}[P_k \leq M|D = 10], d = 5$

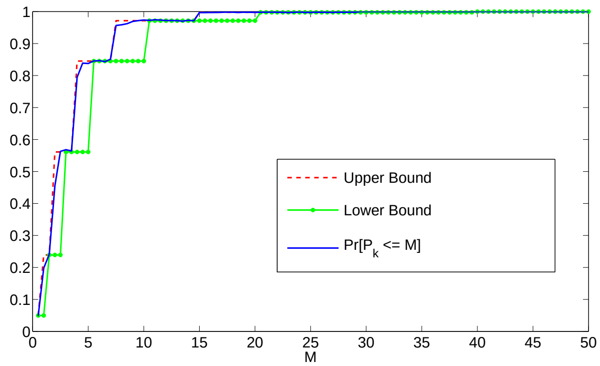


Fig. 9. $\mathbf{Pr}[P_k \leq M|D = 10], d = 3$

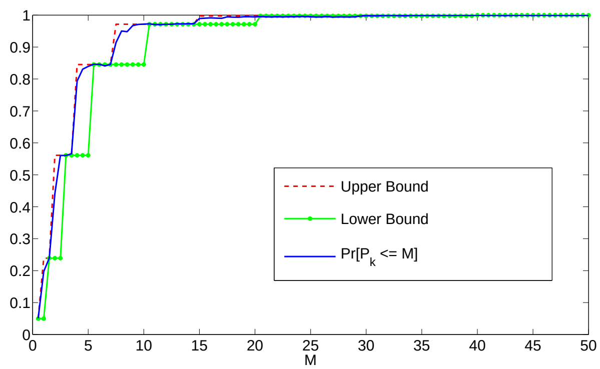


Fig. 10. $\mathbf{Pr}[P_k \leq M|D = 5], d = 3$

$d$ and larger $D$ lead to larger $\mathbf{Pr}[P_k \leq M|D]$, which confirms the theory developed in this chapter. We also notice that when $d = 3$, the filter's performances using $D = 10$ and $D = 5$ only differ slightly (though the former one is better than the latter one), which confirms that using a large buffer may not improve the filter performance drastically.

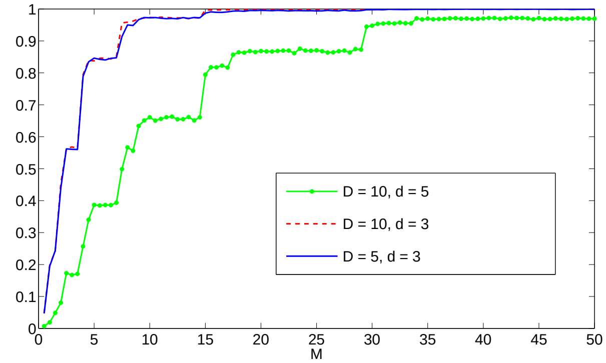


Fig. 11. Comparison of the three simulations

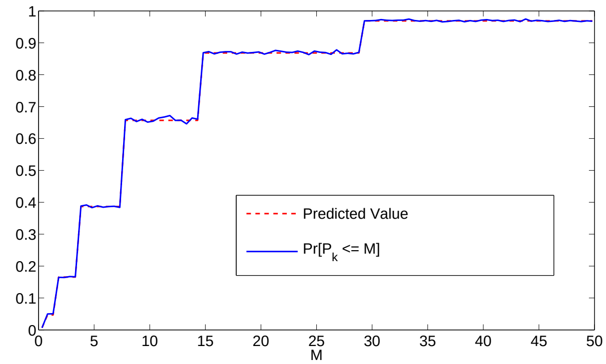


Fig. 12. $\mathbf{Pr}[P_k \leq M|D = 10], d = 5$

### *5.1.2 Sensor with Computation Capability*

We run a Monte Carlo simulation for the case when the sensor has computation capability. Fig. 12 shows the result when $D = 10$ and $d = 5$. As we can see, the predicted value of $\mathbf{Pr}[P_k \leq M|D]$ from Eqn (19) matches well with the actual value.

## *5.2 Vector System*

Consider a vehicle moving in a two dimensional space according to the standard constant acceleration model, which assumes that the vehicle has zero acceleration except for a small perturbation. The state of the vehicle consists of its $x$ and $y$ positions as well as velocities. Assume a sensor measures the positions of the vehicle and sends the measurements to a remote estimator over a packet-delaying network. The system parameters are given according to Eqn (1)-(2) as follows:

$$A = \begin{bmatrix} 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

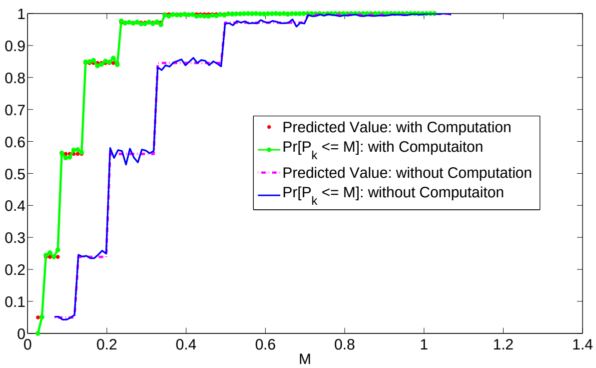


Fig. 13. $\mathbf{Pr}[P_k \leq M|D=5], d=3$

The process and measurement noise covariances are $Q = \mathbf{diag}(0.01, 0.01, 0.01, 0.01)$ and $R = \mathbf{diag}(0.001, 0.001)$. We assume the same delay profile as in the scalar system example with $D = 5$ and $d = 3$.

We run a Monte Carlo simulation for both cases when the sensor has or has not computation capability. As we can see from Fig. 13, the predicted values of $\mathbf{Pr}[P_k \leq M|D]$ from Eqn (19) and Eqn (22) match well with the actual values. We also notice that when sensor has computation capability, the actual filter performance is better than when sensor has no computation capability, as stated in Remark 4.2. In Fig. 13, the $M$ in the x-axis means $M \times I_4$, where $I_4$ is the identity matrix of dimension 4.

## 6 Conclusion

In this paper, we have considered Kalman filtering over a packet-delaying network. Given the distribution of the network induced delay as well as the size of the buffer at the remote estimator, we have characterized the error covariance via a probabilistic approach, i.e., by finding $\mathbf{Pr}[P_k \leq M]$. When measurement data is sent, we give lower and upper bounds on $\mathbf{Pr}[P_k \leq M]$; when estimate data is sent, we provide an exact form on $\mathbf{Pr}[P_k \leq M]$.

There are many interesting work that lie ahead which include: closing the loop using the filtering algorithms proposed in the paper and study closed loop performance; experimentally evaluate the algorithms and theory developed in the paper; extend the results to multi sensor scenarios.

## A Supporting Lemmas

**Lemma A.1** *For any* $0 \leq X \leq Y$,

$$\begin{aligned} &h(X) \leq h(Y), \quad g(X) \leq g(Y), \\ &\tilde{g}(X) \leq \tilde{g}(Y), \quad \tilde{g}(X) \leq X, \\ &h \circ \tilde{g}(X) = g(X), \quad g(X) \leq h(X). \end{aligned}$$

**Proof:** $h(X) \leq h(Y)$ holds as $h(X)$ is affine in $X$. Proof for $g(X) \leq g(Y)$ can be found in Lemma 1-c in [8]. As $\tilde{g}$ is a special form of $g$ by setting $A = I$ and $Q = 0$, we immediately obtain $\tilde{g}(X) \leq \tilde{g}(Y)$. Next we have

$$\tilde{g}(X) = X - XC'[CXC' + R]^{-1}CX \leq X$$

and

$$\begin{aligned} h \circ \tilde{g}(X) &= h(X - XC'[CXC' + R]^{-1}CX) \\ &= A(X - XC'[CXC' + R]^{-1}CX)A' + Q \\ &= g(X). \end{aligned}$$

Finally we have

$$g(X) = h(X) - AXC'[CXC' + R]^{-1}CXA' \leq h(X).$$

**Lemma A.2** *For any* $X \geq 0, \tilde{g}(X) \leq \overline{M}$.

**Proof:** For any $t > 0$, we have

$$\tilde{g}(t\overline{M}) = \frac{t}{t+1}\overline{M} \leq \overline{M}.$$

For all $X \geq 0$, since $\overline{M} > 0$, it is clear that there exists $t_1 > 0$ such that $t_1\overline{M} > X$. Therefore

$$\tilde{g}(X) \leq \tilde{g}(t_1\overline{M}) \leq \overline{M}.$$

**Lemma A.3** $\overline{P} \leq h(\overline{P})$.

**Proof:**

$$h(\overline{P}) = h \circ \tilde{g}(P^*) = g(P^*) = P^* \geq \tilde{g}(P^*) = \overline{P},$$

where the first and the last equality are from the definition of $\overline{P}$, the third equality is from the definition of $P^*$. The rest equality and inequality are from Lemma A.1.

**Lemma A.4** *Let* $X$ *be a continuous random variable defined on* $[0, \infty)$ *and let* $F(x) = \mathbf{Pr}[X \leq x]$. *Then*

$$\mathbb{E}[X] = \int_0^\infty [1 - F(x)]dx.$$

**Proof:** See Lemma (4) in [24], page 93.

## B Evaluate $\mathbb{E}[P_k|D]$ and its Stability

Consider Eqn (1) and (2) with

$$A = a > 1, Q = q > 0, C = c > 0, R = r > 0.$$

For scalar systems, $\mathbf{Pr}[P_k \leq M]$ is the cumulative distribution function of the random variable $P_k$ (for a given $D$). Therefore we can find $\mathbb{E}[P_k]$ from Theorem 3.8 or Theorem 4.1 by using Lemma A.4 in Appendix A, i.e., we write $\mathbb{E}[P_k]$ as

$$\begin{aligned}\mathbb{E}[P_k] &= \int_0^\infty (1-\mathbf{Pr}[P_k \leq M])dM \\ &= \int_0^{\overline{M}} (1-\mathbf{Pr}[P_k \leq M])dM \\ &\quad + \int_{\overline{M}}^\infty (1-\mathbf{Pr}[P_k \leq M])dM.\end{aligned}$$

Let us consider the case when sensor has no computation capability. The following results extends trivially to the case when sensor has computation capability. Using the fact

$$0 \leq \mathbf{Pr}[P_k \leq M] \leq 1,$$

we have

$$\begin{aligned}\mathbb{E}[P_k] &\leq \overline{M} + \int_{\overline{M}}^\infty (1-\mathbf{Pr}[P_k \leq M])dM \\ &\leq \overline{M} + \int_{\overline{M}}^\infty \theta(k_1, D)dM,\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[P_k] &\geq \int_{\overline{M}}^\infty (1-\mathbf{Pr}[P_k \leq M])dM, \\ &\geq \int_{\overline{M}}^\infty \theta(k_2, D)dM.\end{aligned}$$

Recall that $k_1(M) = \min\{t \geq 1 : h^t(\overline{M}) \nleq M\}$ and

$$\begin{aligned}h^t(\overline{M}) &= a^{2t}\overline{M} + q(1 + a^2 + \cdots + a^{2t-2}) \\ &= (\overline{M} + \frac{q}{a^2-1})a^{2t} - \frac{q}{a^2-1} \\ &= c_1 a^{2t} - c_2\end{aligned}$$

where

$$c_1 = \overline{M} + \frac{q}{a^2-1}, c_2 = \frac{q}{a^2-1},$$

therefore for any $t \geq 1$,

$$k_1(M) = t, \text{ if } c_1 a^{2t-2} - c_2 \leq M < c_1 a^{2t} - c_2.$$

Therefore

$$\begin{aligned}\mathbb{E}[P_k] &\leq \overline{M} + \int_{\overline{M}}^\infty \theta(k_1, D)dM \\ &= \overline{M} + \sum_{t=1}^\infty (c_1 a^{2t} - c_1 a^{2t-2})\theta(t, D), \quad \text{(B.1)}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[P_k] &\geq \int_{\overline{M}}^\infty \theta(k_2, D)dM \\ &= \sum_{t=1}^\infty (c_1' a^{2t} - c_1' a^{2t-2})\theta(t, D), \quad \text{(B.2)}\end{aligned}$$

where $c_1' = \overline{P} + \frac{q}{a^2-1}$.

**Lemma B.1** *The minimum $D$ that guarantees*

$$\lim_{k\to\infty} \mathbb{E}[P_k|D] < \infty$$

*is given by*

$$D_{\min} = \min\{d : \sum_{i=0}^{d-1} f(i) > 1 - \frac{1}{a^2}\}. \quad \text{(B.3)}$$

*Any other $D < D_{\min}$ leads to*

$$\lim_{k\to\infty} \mathbb{E}[P_k|D] = \infty.$$

**Proof:** Recall that $\theta(k_i, D)$ is defined in Eqn (9) as

$$\theta(k_i, D) = \prod_{j=0}^{k_i-1} \big(1 - \hat{\gamma}_j(D)\big),$$

and $\hat{\gamma}_i(D) = \hat{\gamma}_D$ for any $i \geq D$. Thus from Eqn (B.1), in order that $\mathbb{E}[P_k|D] < \infty$, it is sufficient that

$$1 - \hat{\gamma}_D < \frac{1}{a^2},$$

or in other words,

$$\sum_{i=0}^{D-1} f(i) > 1 - \frac{1}{a^2}.$$

Similarly, if $D < D_{\min}$, i.e.,

$$1 - \hat{\gamma}_D \geq \frac{1}{a^2},$$

then from Eqn (B.2),

$$\lim_{k\to\infty} \mathbb{E}[P_k|D] = \infty.$$

## References


[1] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, “A survey of recent results in networked control systems,” vol. 95, no. 1. Proceedings of the IEEE, January 2007.

[2] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth-part i: State estimation problems," *IEEE Trans. Automat. Contr.*, vol. 42, Sept 1997.

[3] ——, "Systems with finite communication bandwidth-part ii: Stabilization with limited information feedback," *IEEE Trans. Automat. Contr.*, vol. 44, May 1999.

[4] R. W. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems," *IEEE Trans. Automat. Contr.*, vol. 45, July 2000.

[5] G. N. Nair and R. J. Evans, "Communication-limited stabilization of linear systems," in *Proceedings of the 39th Conf. on Decision and Contr.*, vol. 1, Dec 2000, pp. 1005–1010.

[6] S. C. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[7] I. R. Petersen and A. V. Savkin, "Multi-rate stabilization of multivariable discrete-time linear systems via a limited capacity communication channel," in *Proceedings of the 40th Conf. on Decision and Contr.*, vol. 1, Dec 2001, pp. 304–309.

[8] 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, 2004.

[9] X. Liu and A. Goldsmith, "Kalman filtering with partial observation losses." IEEE Control and Decision, 2004.

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

[11] L. Xie and L. Xie, "Stability of a random riccati equation with markovian binary switching," *IEEE Trans. Automat. Contr., to appear*.

[12] L. Shi, M. Epstein, A. Tiwari, and R. M. Murray, "Estimation with information loss: Asymptotic analysis and error bounds," in *Proceedings of IEEE Conf. on Decision and Control*, Dec 2005, pp. 1215–1221.

[13] A. Ray, L. W. Liou, and J. Shen, "State estimation using randomly delayed measurements," *J. Dyn. Syst., Measurement Contr.*, vol. 115, pp. 19–26, 1993.

[14] E. Yaz and A. Ray, "Linear unbiased state estimation for random models with sensor delay," in *Proceedings of IEEE Conf. on Decision and Control*, Dec 1996, pp. 47–52.

[15] H. Zhang, L. Xie, D. Zhang, and Y. Soh, "A re-organized innovation approach to linear estimation," *IEEE Trans. Automat. Contr., to appear*, vol. 49, no. 10, pp. 1810–1814, 2004.

[16] H. Zhang and L. Xie, *Control and Estimation of Systems with Input/Output Delays*. Springer, 2007.

[17] ——, "Optimal estimation for systems with time-varying delay," in *Proceedings of IEEE Conf. on Decision and Control*, Dec 2007.

[18] A. S. Matveev and A. V. Savkin, "The problem of state estimation via asynchronous communication channels with irregular transmission times," *IEEE Trans. Automat. Contr., to appear*, vol. 48, pp. 670–676, 2003.

[19] S. Nakamori, R. Caballero-Aguila, A. Hermoso-Carazo, , and J. Linares-Perez, "Recursive estimators of signals from measurements with stochastic delays using covariance information," *Applied Mathematics and Computation*, vol. 162, pp. 65–79, 2005.

[20] Z. Wang, D. W. C. Ho, and X. Liu, "Robust filtering under randomly varying sensor delay with variance constraints," *IEEE Trans. Circuits and systems-II: Express briefs*, vol. 51, no. 6, pp. 320–326, 2004.

[21] M. Sahebsara, T. Chen, and S. L. Shah, "Optimal h2 filtering with random sensor delay, multiple packet dropout and uncertain observations," *Int. J of Control*, vol. 80, pp. 292–301, 2007.

[22] S. Sun, L. Xie, W. Xiao, and Y. Soh, "Optimal linear estimation for systems with multiple packet dropouts," *Automatica, to appear*.

[23] L. Shi, M. Epstein, and R. M. Murray, "Kalman filtering over a packet dropping network: a probabilistic approach," in *Tenth International Conference on Control, Automation, Robotics and Vision, Dec 2008, Hanoi, Vietnam. To Appear.*

[24] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford University Press, 2001.