Optimal state estimation over communication channels with random delays

Magdi S. Mahmoud\(^a\),*, Bo Liu\(^b\)

\(^a\) Systems Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
\(^b\) Computer, Electrical and Mathematical Science & Engineering Department, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

Received 24 April 2012; received in revised form 11 November 2012; accepted 19 December 2012
Available online 29 January 2013

Abstract

This paper is concerned with the optimal estimation of linear systems over unreliable communication channels with random delays. The measurements are delivered without time stamp, and the probabilities of time delays are assumed to be known. Since the estimation is time-driven, the actual time delays are converted into virtual time delays among the formulation. The receiver of estimation node stores the sum of arrived measurements between two adjacent processing time instants and also counts the number of arrived measurements. The original linear system is modeled as an extended system with uncertain observation to capture the feature of communication, then the optimal estimation algorithm of systems with uncertain observations is proposed. Additionally, a numerical simulation is presented to show the performance of this work.

© 2013 The Franklin Institute. Published by Elsevier Ltd. All rights reserved.

1. Introduction

It is assumed that the measurements to be processed are transmitted without delay and drop in classical estimation theory [1,2]. However, many applications developed in recent years have incurred unreliable delivery of signals, thus corresponding communication constraints should be considered in system design, for example, finite bandwidth constraints [3], time delay and clock asynchronism constraints [4]. The most popular
application among them is networked control systems (NCSs), i.e., the feedback control systems wherein the control loops are closed through a real-time network (for example, see [5–7]). The major problem associated with NCSs is that measurements arrive at the decision-making location with a nondeterministic delay or could be dropped out totally along the way, therefore it is important to deal with the impact of delay and data packet dropout simultaneously in the overall system performance [8]. Many researchers pay attention to this field, and present some methods, for example, see [9–12,19].

In the most literature, the measurements are encapsulated into packets with time-stamp to indicate the sampling time instant, the decision-making location, or in other words it is called receiver node, obtains easily the exact delivery time delay by subtracting the sampling time instant from the arrival time, see [13–17] for example. However, such strategy is not always possible due to existence of clock asynchronism [4,18] or different networking protocols. In most of existing results of networked control and estimation research, the delay time is calculated by deducting time-stamp from the time at which the packet arrives. Such kind of calculation does not work in practical engineering because of clock asynchronism between two nodes. Since the time delay is random, it is impossible to estimate the clock asynchronism precisely. In this paper, we assume that the data packets delivered are not with time stamp, thus the receiver node cannot get exact time information of the received measurements. However, the probability of time delay is assumed to be known, which sometimes easily obtained from network traffic test statistics. Since the delay time of each measurements is unknown, the process of each received measurements between two adjacent sampling time instant should be same with each others, so the summation of arrived measurements in each time slop is considered instead of every single measurement. The original system is remodeled as a system with uncertain observations to capture the feature of the strategy applied in this work. Since the observation is uncertain, the innovation is not Gaussian distributed, the classical Kalman filter gain cannot be applied for this assumption [1,2]. Some researchers deal with this kind of problem by maximizing the likelihood of corresponding statistics, for example, see [20]. In the paper, a nonlinear filter gain is proposed directly according to the sense of least squares estimation error, thus the optimal estimate of the state is updated by Kalman-filter-liked algorithm. Furthermore, the result obtained in the current work is also suitable for other uncertain observation estimation problems.

The paper is organized as follows. Section 2 presents a linear system and its communication constraints over a unreliable channel. The optimal estimation formulation and algorithm is developed in Section 3. Some useful analysis is proposed in Appendix, and a numerical simulation is presented in Section 4. Some conclusion remarks are given in Section 5.

2. System model and communication constraints

Consider a MIMO discrete time system described in the following state space form:

\[
x_{k+1} = Ax_k + w_k
\]

\[
y_k = Cx_k + v_k
\]

where \(x_k \in \mathbb{R}^n\) and \(y_k \in \mathbb{R}^m\) are the system state and output vectors, respectively. The noise process \(\{w_k\}\) and \(\{v_k\}\) are white, zero-mean, uncorrelated, and have known covariance
matrices $Q$ and $R$. $A$ and $C$ are matrices of appropriate dimensions. Furthermore, it is also assumed that the mathematic expectation of $x_t$ at time instant $t=0$ is $x_0$, then its covariance is $P_0$. Both $x_0$ and $P_0$ are known in advance. Measurements are encapsulated into packets without time-stamp, and then delivered through an unreliable communication channel. It means that the receiver does not have any information of the time when the measurements are delivered.

**Remark 1.** Most networking technology today allows data packets to be time-stamped before they are sent out. It seems like that such time-stamped data packets could benefit the control system analysis and estimation through communication channels, since they supply precise time information on measurements and time delays of delivery. However, in many application cases, for example see [4], the computer clocks located at different network nodes are not synchronized due to the random delivery time delays. In other words, the time stamps in the packets do not make sense for the controller or filter. If the controlled object is linear and the delivery time delay statistic is unknown and bounded, an asynchronous algorithm could be applied. In this paper, we try to solve it by applying alternate method that ignores the time stamps even the data packets sent with time stamps.

The probability density function (PDF) of transmission delay is assumed to be known in advance, i.e., $p(t,k)$ is defined here as probability of an event that a packet is sent at time instant $t$ and it arrives at receiver at time instant $k$. Furthermore, we assume that the probability density function $p(t,k)$ is uniform in $t$, such that $p(t,k) = p(k-t)$.

Note that in most computer-aided engineering practice cases, the digital processors in control systems processes the measurement at every predefined fixed time interval $T$. In other words, it is called time-driven type. If two or more data packets arrive at receiver before the time instant (for example, $iT$) at which the receiver processes them and after the previous processing time instant $(i-1)T$, they are treated as being with same arrival time $iT$. Since the sensors are assumed to work in time-driven type with same frequency, the data packets are treated as being with virtual finite discrete time delays. For example, all data packets with actual transmission delays between $(i-1)T$ and $iT$ are with virtual time delay $(i-1)T$. Let $a_i$ denote the probability of event that a data packet is treated as being with virtual time delay $(i-1)T$, we have

$$a_i = \Pr((i-1)T < \tau_a \leq iT)$$

$$= \int_{(i-1)T}^{iT} p(t) \, dt$$

where $\Pr(\star)$ designates the probability of event “$\star$” and $\tau_a$ designates the actual time delay. It is assumed that the delays of data packets delivery are bounded by a given integer $N$. Let $a_d$ denote the probability of an event that actual time delay is longer than $NT$, i.e.,

$$a_d = \Pr(\tau_a > NT)$$

$$= \int_{NT}^{\infty} p(t) \, dt$$

It is assumed that $a_d = 0$ in this paper.
Remark 2. In much literature if the actual time delay is greater than a predefined number, then the packet is treated as dropout by receiver after its arrival. However, it is not possible to apply a same policy in the current paper, since the measurement data packets are without time stamp, then the receiver cannot decide whether the received packet is processed as dropout packet. The assumption of bounded time delay, giving by Eq. (5), could be almost satisfied by giving a very large N in engineering practice.

Since the delivery delay cannot be smaller than zero, the probability density function $p(t)$ is truncated at the point zero to capture the feature of delivery delays in the same fashion as follows:

$$a_0 = \Pr(\tau_a \leq 0)$$

$$= \int_{-\infty}^{0} p(t) \, dt$$

$$= 0$$

where $a_0$ is defined as probability of an event that the data packet is delivered without time delay.

As for the receiver node, it is assumed that a buffer with two storage units is located at the receiver to store the measurements. One unit of them is a signed floating-point form, which is used to store arrived measurements; The other unit takes up at least enough room for storing an integer, which indicates a counter of number of the arrived data packets between previous processing time instant and the current processing time instant. The interval is $T$, and the round-off errors of the system are assumed to be very trivial and are not considered in the current paper. Fig. 1 shows how to store each arrived measurement. After a some processing time instant, $iT$ for example, both two units are reset to zeros, one of them is used to store the sum of all arrived measurements before next processing time instant $(i+1)T$, the other is used to count how many measurements arrive before time instant $(i+1)T$. At time instant $(i+1)T$, the values in both units are withdrawn by estimator, and both units are reset to zero again.

Since the delivery time delay is bounded by integer $N$, in other words, the potential virtual delay could be in the set $D = \{ \tau_v | \tau_v = 0, T, \ldots, NT \}$, the value of arrived measurements unit at time instant $kT$, which is defined as $z_k$, is sum of one or more of $y_{k-N}, y_{k-N+1}, \ldots, y_k$, i.e.,

$$z_k = d_{k,0}y_k + d_{k,1}y_{k-1} + \cdots + d_{k,N}y_{k-N}$$

where $d_k = [d_{k,0} \, d_{k,1} \, \cdots \, d_{k,N}]^T$ is a random vector, each of its entries is random number that takes value in the set $\{0,1\}$, which indicates if its corresponding measurement arrives.

![Fig. 1. The buffer with two storage units at the receiver side.](image-url)
or not. For instance, \( d_{k,i} = 1 \) implies that measurement \( y(k-i) \) arrives between time interval from \((k-1)T \) and \( kT \); \( d_{k,i} = 0 \) implies that measurement \( y(k-i) \) does not arrive between time interval from \((k-1)T \) and \( kT \). Note that since the data packets are without time stamp, \( d_k \) is totally unknown to the receiver. Furthermore, the value of the other unit designates the number of arrived measurements during the same period, which is defined as \( s_t \) as follows:

\[
s_t = \sum_{j=0}^{N} d_{i,j}
\]

The main goal of the current work is to propose the optimal mean square estimate \( \hat{x}_{k|t}, t \geq k \), its definition is as follows:

\[
\hat{x}_{k|t} = \mathbb{E}\{x_k|z', S', x_0, P_0\}, \quad t \geq k
\]

where symbol \( \mathbb{E}\{\cdot\} \) denotes the mathematic expectation, \( z' \) is shorthand notation of received measurements \( \{z_0, z_1, \ldots, z_t\} \), and \( S' \) is defined as historical data of \( s_i \) up to time instant \( tT \), i.e., \( S' = \{s_0, s_1, \ldots, s_t\} \). The estimation error is defined as \( e_{k|t} = x_k - \hat{x}_{k|t} \) and error covariance \( P_{k|t} = \mathbb{E}\{e_{k|t}e_{k|t}^T|z', S', x_0, P_0\} \).

3. Optimal estimation algorithm

3.1. System augmentation and preparatory formulation

According to the assumptions mentioned at previous section, modeling system (1)–(2) by considering communication channel feature yields

\[
\xi_{k+1} = F \xi_k + w_k
\]

\[
z_k = H_k \xi_k + E_k v_k
\]

\[
s_k = \sum_{j=0}^{N} d_{k,j}
\]

where \( \xi_k = [x_k^T \ x_{k-1}^T \ \cdots \ x_{k-N}^T]^T, \ w_k = [w_k^T \ 0 \ \cdots \ 0]^T, \ v_k = [v_k^T \ u_{k-1}^T \ \cdots \ u_{k-N}^T]^T \), and constant system matrix \( F \) is obtained from system (1)–(2) as follows:

\[
F = \begin{bmatrix}
A & 0 & 0 & \cdots & 0 & 0 \\
I_n & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & I_n & 0
\end{bmatrix}
\]

and \( H_k \in \{H^i, i = 0,1, \ldots, 2^{N+1}-1\} \) and \( E_k \in \{E^i, i = 0, \ldots, 2^{N+1}-1\} \). By expressing \( i \) in binary number

\[
i = d_{k,0}d_{k,1} \cdots d_{k,N}B = \sum_{j=0}^{N} d_{k,j}2^{N-j}
\]
where “\(\ast B\)” denotes that “\(\ast\)” is a binary number, and \(d_{ij} \in \{0,1\}, j = 0,2,\ldots,N\) are defined at previous section, thus the matrices \(H_k\) and \(E_k\) are defined as follows:

\[
H_k = [d_{k,0}C \ d_{k,1}C \cdots \ d_{k,N}C]
\]

\[
E_k = [d_{k,0}I_m \ d_{k,1}I_m \cdots \ d_{k,N}I_m]
\]

and the definitions of \(H^i\) and \(E^i\) are as follows:

\[
H^i = [a_0C \ a_1C \cdots \ a_NC], \quad a_0a_1 \cdots a_NB = i
\]

\[
E^i = [a_0I_m \ a_1I_m \cdots \ a_NI_m], \quad a_0a_1 \cdots a_NB = i
\]

for example, if \(N=2, d_{k,0} = 1, d_{k,1} = 0\) and \(d_{k,2} = 1, i = 101B = 4 + 0 + 1 = 5\), and the corresponding \(H_2 = H^5 = [C \ 0 \ C]\), this is a kind of notation to construct a one-to-one relationship between the elements in a matrix set and the elements of 10-based number (or 2-based number).

Before we present the optimal estimation algorithm, we have to introduce some preparatory work. Table 1 is an array of indicators \(d_{n,r}\), and the sum of all elements in each row of the array is its corresponding \(s_n\), for example, the sum of the elements in the first row, i.e., \(d_{k,N} + d_{k,N-1} + \cdots + d_{k,0}\) is \(s_k\). Since the measurement \(y_k\) could arrive at the receiver node once and only once, the sum of all elements in each column of the array is one, for example, \(d_{k,N} + d_{k-1,N-1} + \cdots + d_{k-N,0} = 1\). According to the analysis above, we can conclude that the maximum of \(s_k\) based on the knowledge of \(S^{k-1}\) is as follows:

\[
s_{k|k-1}^{\text{max}} \triangleq \max\{s_k|S^{k-1}\} = 1 + N - \sum_{j=1,s_{k-j} \geq j}^{N} (s_{s-j} - j)
\]

where the term 1 at the right side of Eq. (22) refers to the fact that \(Pr[d_{k,0} = 1]\) is always greater than zero, and the last term refers to the fact that as for each \(s_{k-j}\), there are at least \((s_{s-j} - j)\) indicators in the same row affect the value of \(s_k\). Take the second row of the array as an example, \(d_{k-1,N}\) is not located in the same column with any indicators of first row, and the other \(N-1\) indicators, i.e., from \(d_{k-1,N-1}\) to \(d_{k-1,0}\), lay in the same column with the indicators from \(d_{k,N}\) to \(d_{k,1}\), respectively. Thus, if \(s_{k-1} \geq 1\), there at least \(s_{k-1} - 1\) should be deducted from the maximum of the \(s_k\).

Now, define a set of possible observation matrix at time instant \(tT\) based on the knowledge of \(S^{k-1}\) as follows:

\[
\Omega_{k|k-1}^{\text{p}} \triangleq \left\{H^i|i = d_0d_1 \cdots d_NB, \sum_{j=0}^{N} d_j \leq s_{i|k-1}^{\text{max}}\right\}
\]

Table 1
The number of arrived measurements up time instant \(kT\).

<table>
<thead>
<tr>
<th>(d_{k,N})</th>
<th>(d_{k-1,N-1})</th>
<th>(d_{k-2,N-2})</th>
<th>(d_{k-3,N-3})</th>
<th>(y_k-N)</th>
<th>(y_k-N)</th>
<th>(y_k-N)</th>
<th>(y_k-N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_{k-1,N})</td>
<td>(d_{k-1,2})</td>
<td>(d_{k-1,1})</td>
<td>(d_{k-1,0})</td>
<td>(s_k)</td>
<td>(s_k)</td>
<td>(s_k)</td>
<td>(s_k)</td>
</tr>
<tr>
<td>(d_{k-2,N})</td>
<td>(d_{k-2,1})</td>
<td>(d_{k-2,0})</td>
<td>(s_k)</td>
<td>(s_k)</td>
<td>(s_k)</td>
<td>(s_k)</td>
<td>(s_k)</td>
</tr>
<tr>
<td>(d_{k-3,N})</td>
<td>(d_{k-3,0})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1
The number of arrived measurements up time instant \(kT\).
Especially, if $\delta_{\text{max}}^{\text{t}} = 1$, then
\begin{equation}
\Omega^o_{k|k-1} \triangleq \{H^i|d_0 d_1 \cdots B_0 = 0\text{ or } 1\}
\end{equation}

it is because at any time instant $k$, $\Pr[d_k, 0 = 1]$ is always greater than zero.

The estimate of $H_t$ with knowledge of $S^{t-1}$ is obtained by
\begin{equation}
\hat{H}_{k|k-1} \triangleq E\{H_k|S^{k-1}\}
\end{equation}

where the notation $C^n_m$ indicates the combination of $n$ from $m$,
$q(i) = \sum_{j=0}^{N}d_j$ when expressing $i = d_0 d_1 \cdots d_N B$.
\begin{equation}
\Pr[H_k = H^i] \text{ is easily obtained by}
\end{equation}
\begin{equation}
\Pr[H_k = H^i], \forall k
\end{equation}
\begin{equation}
= \prod_{j=0}^{N}(d_{i,j}\Pr[\tau_v = j] + (1 - d_{i,j})\Pr[\tau_v \neq j])
\end{equation}
\begin{equation}
= \prod_{j=0}^{N}((2d_{i,j} - 1)\Pr[\tau_v = j] + 1 - d_{i,j})
\end{equation}

Remark 3. To avoid confusion, one should note that even $H^i$ and $H_k$ are with same size,
but $H^i$ is a particular element in set $\Omega^o_{k|k-1}$ and it is determinant; $H_k$ is random matrix
depends on time index $k$, and moreover, $H_k$ must be equal to one element of in set $\Omega^o_{k|k-1}$.

3.2. Optimal estimation of the augmented system

Define the optimal estimate of $\xi_t$ at time instant $tT$ as
\begin{equation}
\hat{\xi}_{t|t} \triangleq E\{\xi_t|z^t, S^t, x_0, P_0\}
\end{equation}
such that $\hat{x}_{k|t} = \hat{\xi}_{(t-k+1)|t}$, where $\hat{\xi}_{(t-k+1)|t}$ designates the $(t-k+1)$th block element of $\hat{\xi}_t$ with
dimension $n$. Rewrite the optimal estimate formula as follows:
\begin{equation}
\hat{\xi}_{t|t} = E\{\xi_t|z^t, S^t, x_0, P_0\}
\end{equation}
\begin{equation}
= E\{\hat{\xi}_t|z^t-1, z_t, S^{t-1}, s_t, x_0, P_0\}
\end{equation}
\begin{equation}
= E\{\xi_t|z^t-1, S^{t-1}, x_0, P_0\} + E\{\hat{\xi}_{t|t-1}|z_{t|t-1}, s_{t|t-1}\}
where
\[
\hat{\xi}_{t|t-1} = \hat{\xi}_t - \mathbb{E}\{\xi_t|z^{t-1}, S^{t-1}, x_0, P_0\}
\]
(36)
denotes a priori estimation error and the pair
\[
(\hat{\xi}_{t|t-1}, \hat{s}_{t|t-1}) = (z_t, s_t) - \mathbb{E}\{(z_t, s_t)|z^{t-1}, S^{t-1}, x_0, P_0\}
\]
(37)
Note that \(\hat{\xi}_{t|t-1}\) and \(\hat{s}_{t|t-1}\) are presented as two independent variables before, however, according to the formulation in the following part of the current paper, the information of \(\hat{s}_{t|t-1}\) is included in variable \(\hat{\xi}_{t|t-1}\), we use \(z_t\) and \(\hat{s}_{t|t-1}\) instead of pair \((z_t, s_t)\) and \((\hat{\xi}_{t|t-1}, \hat{s}_{t|t-1})\) in the rest part of this paper.

The first term at the right side of Eq. (35) is
\[
\mathbb{E}\{\xi_t|z^{t-1}, S^{t-1}, x_0, P_0\} = \mathbb{E}\{F\hat{\xi}_{t-1} - \sigma_{t-1}^{1/2}\}z^{t-1}, S^{t-1}, x_0, P_0\}
\]
(38)
\[
= \mathbb{E}\{F\hat{\xi}_{t-1} + \sigma_{t-1}^{1/2}\}z^{t-1}, S^{t-1}, x_0, P_0\}
\]
(39)
\[
= \mathbb{E}\{F\hat{\xi}_{t-1}|z^{t-1}, S^{t-1}, x_0, P_0\} + \mathbb{E}\{\sigma_{t-1}^{1/2}|z^{t-1}, S^{t-1}, x_0, P_0\}
\]
(40)
\[
= \tilde{F}\hat{\xi}_{t-1|t-1}
\]
(41)
thus Eq. (36) becomes
\[
\hat{\xi}_{t|t-1} = \hat{\xi}_t - \tilde{F}\hat{\xi}_{t-1}
\]
(42)
\[
= F\hat{\xi}_{t-1} + \sigma_{t-1}^{1/2} - \tilde{F}\hat{\xi}_{t-1}
\]
(43)
\[
= F\hat{\xi}_{t-1|t-1} + \sigma_{t-1}^{1/2}
\]
(44)
where \(\xi_{t-1|t-1} = \hat{\xi}_{t-1} - \hat{\xi}_{t-1|t-1}\), the recursive update of a priori estimation error covariance is as follows:
\[
\Sigma_{t|t-1} \triangleq \text{cov} [\hat{\xi}_{t|t-1}, \hat{\xi}_{t|t-1}] = F\Sigma_{t-1|t-1}F^T + \bar{Q}
\]
(45)
where \(\Sigma_{t-1|t-1} \triangleq \text{cov} [\hat{\xi}_{t-1|t-1}, \hat{\xi}_{t-1|t-1}]\) is a posteriori estimation error covariance, and \(\bar{Q} = \text{diag}(Q, 0, 0, \ldots, 0)\) we have
\[
\tilde{z}_{t|t-1} = z_t - \hat{z}_{t|t-1}
\]
(46)
\[
= H_t\hat{\xi}_t + E_t \bar{v}_t - \mathbb{E}\{z_t|z^{t-1}, S^{t-1}, x_0, P_0\}
\]
(47)
\[
= H_t\hat{\xi}_t + E_t \bar{v}_t - \mathbb{E}\{H_t\hat{\xi}_t|z^{t-1}, S^{t-1}, x_0, P_0\}
\]
(48)
\[
= H_t\hat{\xi}_t + E_t \bar{v}_t - \mathbb{E}\{H_t|z^{t-1}, S^{t-1}, x_0, P_0\}\mathbb{E}\{\xi_t|z^{t-1}, S^{t-1}, x_0, P_0\}
\]
(49)
\[
= H_t\tilde{z}_t + E_t \bar{v}_t - \mathbb{E}\{H_t|z^{t-1}, S^{t-1}, x_0, P_0\}\hat{\xi}_{t|t-1}
\]
(50)
\[
= H_t\tilde{z}_t - \hat{H}_{t|t-1}\hat{\xi}_{t|t-1} + E_t \bar{v}_t
\]
(51)
\[
= H_t\tilde{z}_t - (H_t - \hat{H}_{t|t-1})\hat{\xi}_{t|t-1} + E_t \bar{v}_t
\]
(52)
Let $M^i_t = [H^i_t (\hat{H}_t^{i-1} - H^i_t) \ E^i_t]$ and $p^i = [\tilde{\xi}^T_t \ \tilde{\bar{\xi}}^T_{t-1} \ \tilde{v}^T_t]^T$, thus the last term of Eq. (35) is formulated as follows:

$$
E\{\tilde{\xi}_t^{i-1} | \tilde{z}_t^{i-1}, \tilde{\bar{\xi}}_t^{i-1}\} = E\{\tilde{\xi}_t^{i-1} | M_i \phi_t\}
$$

(53)

where $M_t \in \{M^i_t | i = 1, 2, \ldots \}$ is random matrix and the $p^i_M = \text{Pr} \{M_t = M^i\} = \text{Pr} \{H_t = H^i\}$.

$H^i_t \in \Omega^p_{t-1}$ is obtained. Since $\text{cov} \{\tilde{\xi}_t^{i-1}\} = \Sigma_{t-1}^{i-1}$, $P^i_{\tilde{\xi}_t^{i-1}, \phi_t} = [\Sigma_{t-1} 0 0] = P^T_{\phi_t, \tilde{\xi}_t^{i-1}}$ and $P_{\phi_t} = \text{diag} \{\Sigma_{t-1}, \tilde{\Sigma}_{t-1}, \tilde{\bar{R}}_t\}$, where $\tilde{\Sigma}_{t}^{i-1} = E \{\tilde{\xi}_t^{i-1} \tilde{\xi}_t^{i-1}^T\} = \Theta_t + \Sigma_{t-1}^{i-1}$, since $\Theta_t$ is actually the covariance of $\tilde{\xi}_t$, the recursive updating of $\Theta_t$ is easily obtained based on the knowledge of $P_0$ and the system model (15), i.e., $\Theta_t = F \Theta_{t-1} F^T + Q_t$, the initial value $P_0 = \text{diag} \{P_0, P_1, \ldots, P_N\}$, $P_j = A P_{j-1} A^T + Q_j$ for $j = 1, 2, \ldots, N$ and $\bar{R}_t = \text{diag} \{R, R, \ldots, R\}$. Applying the algorithm in Theorem 1 yields

$$
E\{\tilde{\xi}_t^{i-1} | \tilde{z}_t^{i-1}\} = \sum_{H_t \in \Omega^p_{t-1}} K^i_t \tilde{z}_t^{i-1} - (\hat{H}_t^{i-1} - H^i_t) \tilde{\bar{\xi}}_t^{i-1}
$$

(54)

where the filter gain $K^i_t = \gamma^i_t B^i_t$, and where the coefficients

$$
\gamma^i_t = \frac{p^i_M^i | M^i_t P^i_t M^T_t | -1/2 \exp \left\{ -\frac{1}{2} \tilde{\xi}_t^{i-1} \left[ M^i_t P^i_t M^T_t \right]^{-1} \tilde{\xi}_t^{i-1} \right\}}{\sum_{H_t \in \Omega^p_{t-1}} p^i_M^i | M^i_t P^i_t M^T_t | -1/2 \exp \left\{ -\frac{1}{2} \tilde{\xi}_t^{i-1} \left[ M^i_t P^i_t M^T_t \right]^{-1} \tilde{\xi}_t^{i-1} \right\}}
$$

(55)

where $\tilde{\xi}_t^{i-1} = \tilde{z}_t^{i-1} - (\hat{H}_t^{i-1} - H^i_t) \tilde{\bar{\xi}}_t^{i-1}$ and the matrices $B^i_t = -P^i_{\tilde{\xi}_t^{i-1}, \phi_t} M^T_t [M^i_t P^i_t M^T_t]^{-1}$. The conditional estimation error covariance is as follows:

$$
\Gamma_t = \sum_{H_t \in \Omega^p_{t-1}} \gamma^2_t \Gamma^i_t
$$

where $\Gamma^i_t = \Sigma_{t-1} - P^i_{\tilde{\xi}_t^{i-1}, \phi_t} M^T_t [M^i_t P^i_t M^T_t]^{-1} M^i_t P^i_t$, $\tilde{\xi}_t^{i-1}$.

Since we have the following formulation of a posterior estimation error covariance:

$$
\tilde{\xi}_t^{i-1} \equiv \tilde{\xi}_t^{i-1} - \tilde{\bar{\xi}}_t^{i-1}
$$

(56)

$$
= \tilde{\xi}_t^{i-1} - E\{\tilde{\xi}_t^{i-1} | \tilde{z}_t^{i-1}\}
$$

(57)

$$
= \tilde{\xi}_t^{i-1} - \sum_{H_t \in \Omega^p_{t-1}} K^i_t \xi_t^{i-1}
$$

(58)

thus $\Sigma_{t|t}$, the covariance matrix of estimation error $\tilde{\xi}_t^{i-1}$ is obtained as follows:

$$
\Sigma_{t|t} = \text{cov} \{\tilde{\xi}_t^{i-1}, \tilde{\xi}_t^{i-1}\} - \sum_{H_t \in \Omega^p_{t-1}} K^i_t \text{cov} \{\xi_t^{i-1}, \xi_t^{i-1}\} K^i_T
$$

(59)

$$
= \Sigma_{t|t-1} - \sum_{H_t \in \Omega^p_{t-1}} K^i_t [H_t|t-1 \Sigma_{t-1} H^T_t] + E_{t-1} \bar{Q} E^T_{t-1}
$$

(60)

$$
+ (\hat{H}_t^{i-1} - H^i_t) \tilde{\Sigma}_{t-1}^{i-1} (\hat{H}_t^{i-1} - H^i_t)^T K^i_T
$$

(61)
Finally, the optimal estimate $\hat{x}_{t|t}$ of the extended system (35) is solved, the corresponding optimal estimate of the state $x_t$ could be picked out from $\hat{x}_{t|t}$, with same fashion, the estimation error of $\hat{x}_{t|t}$ is also available in $\Sigma_{t|t}$.

4. Numerical simulations

A simple numerical example is applied here to show the performance of the proposed method in this section. System (1)–(2) is evaluated as follows:

$$A = \begin{bmatrix} 0.71 & 0.82 \\ 0 & 0.2 \end{bmatrix}, \quad C = [0.8 \ 0.1]$$

(62)

and the noise variance matrices are

$$Q = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.5 \end{bmatrix}, \quad R = 0.2$$

(63)

The actual random time delay is assumed as $\tau_a \sim N(5T, T^2)$, the maximum delay is truncated as $N=10$, where $T$ denotes the sampling step size and since the system is given in discrete-time from, $T$ refers to one in this case. Fig. 2 shows the measurements $y_t$ and the number of received measurements at each step from a trial of 200 steps length computer simulation. The system state and their corresponding estimate are shown in Fig. 3, wherein the solid lines denote the true states, $x^1$ in the upper figure, and $x^2$ in the bottom figure. The dark dash lines denote the state estimates, the light dash lines denote the 4 steps backward smoothed estimates and the darker dash lines denote the 8 steps backward smoothed estimates. The absolute of errors of both filtering and smoothing is shown in Fig. 4. The upper figure of Fig. 4 shows the absolute of filtering errors $|e_{t|t}|$, where the red dash line refers to the first element of $|e_{t|t}|$, and the blue solid line refers to the second element. Since there are 10 smoothing errors involved in this simulation, i.e., from $e_{t|t+1}$ to $e_{t|t+N}$, we just pick up two of them as examples for short. The middle figure and the bottom figure of Fig. 4 show the curves of $e_{t|t+4}$ and $e_{t|t+8}$, respectively. With the same fashion, the red dash

![Fig. 2. The measurements $y_t$ and the number of received measurements $s_t$.](image-url)
lines refer to the first element of corresponding vector, and the blue solid lines refer to the second element. Visual inspection of each subplotation reveals a strong trend that the greater backstep size $b$ is, the smaller the smoothing error $|e_{t \| t+b} \rangle$ is.
5. Conclusions

The optimal estimation of linear systems over unreliable communication channels based on delayed measurements is proposed in this work. Since the received measurements are assumed to be without time stamps, the linear system considered in this paper is remodeled as an augmented system with uncertain observations. The received measurements between two adjacent processing time instant are summed up, and the summation is processed with an additional information, the number of summed measurements. Hence an optimal estimation method for systems with uncertain observations is developed, and the optimal filter gain turns up to be nonlinear function of measurements. The technique developed in this work is also suitable for optimal estimation problems in other fields. The main contribution of this paper is twofold, first of all, the proposed estimation formulation does not depend on the time-stamp, which makes the method practical in engineering and, secondly, the optimal estimation under uncertain observations has strong potential to be extended to other applications. Finally, a numerical example is applied to show the performance of the proposed method.

Acknowledgment

The authors would like to thank the reviewers for their careful assessments and constructive comments on our submission. This work is supported by the deanship for scientific research (DSR) at KFUPM through group research project RG-1105-1.

Appendix

Lemma 1. Consider a n-dimensional Gaussian random vector $X \sim N(m_X, P_X)$ and a random linear function of $X$: $Y = AX$, where $A \in \{A_i \in R^{n \times n} | i = 1,2, \ldots, N\}$ is a random matrix, which is uncorrelated with the random vector $X$ and the probability of each $A_i$, i.e., $p_A^i \triangleq \Pr(A = A_i), \forall h = 1,2, \ldots, N \text{ is known},$ then the mathematic expectation of the random vector $Y$ is $\bar{A}m_Y$ and its variance is $\bar{A}P_X\bar{A}^T$, where $\bar{A} \triangleq \sum_{i=1}^{N} p_A^i A_i$. Furthermore, the PDF of random vector $Y$ is as follows:

$$f_Y(y) = \sum_{i} p_A^i (2\pi)^{-m/2} |A_i P_X A_i^T|^{-1/2} \exp\left\{-\frac{1}{2}(y-A_i m_X)^T (A_i P_X A_i^T)^{-1} (y-A_i m_X)\right\}$$

(64)

Proof. Since the random matrix $A$ is uncorrelated with the Gaussian random vector $X$, we have the following formulation:

$$m_Y \triangleq E[Y]$$

(65)

$$= E[A_i X]$$

(66)

$$= E[A_i] E[X]$$

(67)

$$= \sum_{i=1}^{N} p_A^i A_i m_X$$

(68)
$$= \bar{A}m_X$$

(69)

and

$$P_Y \triangleq E[(Y - E(Y))(Y - E(Y))^T]$$

(70)

$$= E[YY^T - Ym_Y^T - m_Y Y^T + m_Y m_Y^T]$$

(71)

$$= \bar{A}E[XX^T - Xm_X^T - m_X X^T + m_X m_X^T] \bar{A}^T$$

(72)

$$= \bar{A}P_X \bar{A}^T$$

(73)

$$\Pr(Y \leq y)$$

(74)

$$= \Pr(AX \leq y)$$

(75)

$$= \sum_i \Pr(AX \leq y | A = A_i) \Pr(A = A_i)$$

(76)

$$= \sum_i \Pr(A_iX \leq y) \Pr(A = A_i)$$

(77)

$$= \sum_i p_A(i) \Pr(A_iX \leq y)$$

(78)

Since the PDF of Gaussian random $X$ is

$$f_X(x) = (2\pi)^{-n/2} |P_X|^{-1/2} \exp\left\{- \frac{1}{2}(x - m_X)^T P_X^{-1} (x - m_X) \right\}$$

and for any nonrandom matrix $A$, the PDF of $A_iX$ is

$$f_{A_iX}(y) = (2\pi)^{-m/2} |A_iP_X A_i^T|^{-1/2} \exp\left\{- \frac{1}{2}(y - A_i m_X)^T (A_i P_X A_i^T)^{-1} (y - A_i m_X) \right\}$$

then the PDF of random vector $Y$ is

$$f_Y(y)$$

(79)

$$= \frac{d}{dy} \Pr(Y \leq y)$$

(80)

$$= \frac{d}{dy} \left( \sum_i p_A^i \Pr(A_iX \leq y) \right)$$

(81)

$$= \sum_i p_A^i (2\pi)^{-m/2} |A_i P_X A_i^T|^{-1/2} \exp\left\{- \frac{1}{2}(y - A_i m_X)^T (A_i P_X A_i^T)^{-1} (y - A_i m_X) \right\}$$

(82)

**Lemma 2.** Consider a set of independent Gaussian random vectors with same dimension $X_i \in \mathbb{R}^n \sim N(m_i, P_i)$, $i = 1, 2, \ldots, N$. Assume that $Y$ is random observation of these $X_i$ with known probabilities, i.e., $p_i \triangleq \Pr(Y = X_i)$, $i = 1, 2, \ldots, N$ is known and $\sum_{i=1}^{N} \Pr(Y = X_i) = 1$, then the mathematic expectation and variance of $Y$ is $\sum_{i=1}^{N} p_i m_i$ and $\sum_{i=1}^{N} p_i^2 P_i$. 


respectively. Furthermore, the PDF of $Y$ is

$$f_Y(y) = (2\pi)^{-n/2} \sum_{i=1}^{N} p_i |P_i|^{-1/2} \exp\left\{-\frac{1}{2} (y-m_i)^T (P_i)^{-1} (y-m_i)\right\}$$  \(83\)

**Proof.** Define a random matrix $A \in \mathbb{R}^{n \times n}$, and $\Pr(A = A_i) = \Pr(Y = X_i), i = 1, 2, \ldots, N$, where $A_i$ is given as follows:

$$A_1 = [I_n \ 0 \ 0 \ \cdots \ 0]$$  \(84\)

$$A_2 = [0 \ I_n \ 0 \ \cdots \ 0]$$  \(85\)

$$\vdots$$  \(86\)

$$A_N = [0 \ 0 \ 0 \ \cdots \ I_n]$$  \(87\)

Define that

$$\overline{A} = \sum_{i=1}^{N} p_i A_i$$  \(88\)

$$= [p_1 I_n \ p_2 I_n \ p_3 I_n \ \cdots \ p_N I_n]$$  \(89\)

Rearranging $X_i$ as

$$A = [X_1^T \ X_2^T \ X_3^T \ \cdots \ X_N^T]^T$$  \(90\)

the mathematic expectation of $A$ is

$$E[A] \triangleq m = [m_1^T \ m_2^T \ m_3^T \ \cdots \ m_N^T]^T$$  \(91\)

Since $X_i$ is independent to $X_j, i \neq j$, the variance of random vector $A$ is

$$P_A = \text{diag}[P_1 \ P_2 \ P_3 \ \cdots \ P_N]$$

thus the random observation $Y$ is represented as $Y = AA$. According to the result of Lemma 1, the mathematic expectation of $Y$ is $E[Y] = \overline{A}m = \sum_{i=1}^{N} p_im_i$ and the variance is as follows:

$$P_Y = \overline{A} P_A \overline{A}^T$$  \(92\)

$$= \sum_{i=1}^{N} p_i^2 P_i$$  \(93\)

Furthermore, the PDF of $Y$ is given by

$$f_Y(y) = \frac{d}{dy} \Pr\{Y \leq y\}$$  \(94\)

$$= \frac{d}{dy} \Pr\{AA \leq y\}$$  \(95\)

$$= \sum_{i=1}^{N} \frac{d}{dy} \Pr\{AA \leq y|A = A_i\} \Pr\{A = A_i\}$$  \(96\)
\[\begin{align*}
&= \sum_{i=1}^{N} p_i \frac{d}{dy} \Pr[X_i \leq y] \\
&= (2\pi)^{-n/2} \sum_{i=1}^{N} p_i |P_i|^{-1/2} \exp \left\{-\frac{1}{2} (y-m_i)^T (P_i)^{-1} (y-m_i) \right\} \quad \square
\end{align*}\]

**Theorem 1.** Consider two jointly distributed Gaussian random vector \( X \sim N(m_X, P_X) \) and \( Y \sim N(m_Y, P_Y) \) with respective dimensions \( n \) and \( l \), whose covariance is \( P_{XY} \) and a random vector \( W = AY \), where \( A \in \mathbb{R}^{n \times l} \) \( i = 1, 2, \ldots, N \) is a random matrix and uncorrelated with the random vectors \( X \) and \( Y \). The probability of each \( A_i \), i.e., \( p_i \triangleq \Pr[A = A_i] \), \( \forall h = 1, 2, \ldots, N \) is known, and also suppose that in a particular sample observation the value of \( W \) is measured to be \( w \), then the optimal estimate (in the sense of least-squared estimation error) of \( X \) of the corresponding sample value \( w \) is

\[\hat{x} \triangleq \mathbb{E}[X \mid W = w] = \sum_{i=1}^{N} \gamma_i B_i \tilde{w}_i\]

where the coefficients

\[\gamma_i = \frac{p_i |A_i P_Y A_i^T|^{-1/2} \exp \left\{-\frac{1}{2} w^T (A_i P_Y A_i^T)^{-1} w \right\}}{\sum_{i=1}^{N} p_i |A_i P_Y A_i^T|^{-1/2} \exp \left\{-\frac{1}{2} w^T (A_i P_Y A_i^T)^{-1} w \right\}}\]

for \( i = 1, 2, \ldots, N \), \( \tilde{w}_i = w - A_i m_X \), and the matrices \( B_i = -P_{XY} A_i^T (A_i P_Y A_i^T)^{-1} \).

The estimation error covariance is as follows:

\[P \triangleq \mathbb{E}[\|X - \hat{x}\|^2 \mid W = w] \]

\[= \sum_{i=1}^{N} \gamma_i^2 P_i\]

where \( P_i = P_X - P_{XY} A_i^T (A_i P_Y A_i^T)^{-1} A_i P_{YX} \), for \( i = 1, 2, \ldots, N \).

**Proof.** Let \( Z = [X^T \ Y^T]^T \), and random vector \( Z \) can be written as \( Z = \tilde{A} [X^T \ Y^T]^T \), where

\[\tilde{A} = \begin{bmatrix}
I_n & 0 \\
0 & A
\end{bmatrix}\]

we also define that

\[\overline{A} \triangleq \sum_{i} p_i \begin{bmatrix}
I_n & 0 \\
0 & \overline{A}
\end{bmatrix}\]

\[= \begin{bmatrix}
I_n & 0 \\
0 & \overline{A}
\end{bmatrix}\]

we have

\[\mathbb{E}[Z] = \begin{bmatrix}
m_X \\
\overline{A} m_Y
\end{bmatrix}\]
and covariance

$$P_Z = \begin{bmatrix} P_X & P_{XY} A^T \\ \overline{AP}_Y & \overline{AP}_Y A^T \end{bmatrix}$$

$$= \begin{bmatrix} I_n & 0 \\ 0 & \overline{A} \end{bmatrix} \begin{bmatrix} P_X & P_{XY} \\ P_{YX} & P_Y \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & \overline{A} \end{bmatrix}^T$$

then the joint PDF of random vectors $X$ and $W$ is

$$f_{X,W}(x,w) = f_Z(z)$$

$$= \frac{d}{dz} \Pr(Z \leq z)$$

$$= \frac{d}{dz} \left( \sum_i p_i \Pr(\tilde{A}_i Z \leq z) \right)$$

$$= \sum_i p_i (2\pi)^{-m/2} |\tilde{A}_i P_Z \tilde{A}_i^T|^{-1/2}$$

$$\exp\left\{ -\frac{1}{2} (z-\tilde{A}_i m_Z)^T (\tilde{A}_i P_Z \tilde{A}_i^T)^{-1} (z-\tilde{A}_i m_Z) \right\}$$

The conditional density of $X$ given $W$ is

$$f_{X|W}(x|w) = f_{X,W}(x,w)[f_W(w)]^{-1}$$

$$= (2\pi)^{-n/2} \frac{\sum_i p_i |\tilde{A}_i P_Z \tilde{A}_i^T|^{-1/2} \exp\left\{ -\frac{1}{2} (z-\tilde{A}_i m_Z)^T (\tilde{A}_i P_Z \tilde{A}_i^T)^{-1} (z-\tilde{A}_i m_Z) \right\}}{\sum_i p_i |A_i P_X A_i^T|^{-1/2} \exp\left\{ -\frac{1}{2} (w-A_i m_X)^T (A_i P_Y A_i^T)^{-1} (w-A_i m_X) \right\}}$$

Substituting $\tilde{z}_i = z-\tilde{A}_i m_Z$ and $\tilde{w}_i = w-A_i m_X$ yields

$$f_{X|W}(x|w)$$

$$= f_{X,W}(x,w)[f_W(w)]^{-1}$$

$$= (2\pi)^{-n/2} \frac{\sum_i p_i |\tilde{A}_i P_Z \tilde{A}_i^T|^{-1/2} \exp\left\{ -\frac{1}{2} \tilde{z}_i^T (\tilde{A}_i P_Z \tilde{A}_i^T)^{-1} \tilde{z}_i \right\}}{\sum_i p_i |A_i P_X A_i^T|^{-1/2} \exp\left\{ -\frac{1}{2} \tilde{w}_i^T (A_i P_Y A_i^T)^{-1} \tilde{w}_i \right\}}$$

Let

$$a_i = p_i |\tilde{A}_i P_Z \tilde{A}_i^T|^{-1/2} \exp\left\{ -\frac{1}{2} \tilde{z}_i^T (\tilde{A}_i P_Z \tilde{A}_i^T)^{-1} \tilde{z}_i \right\}$$

$$\beta_i = p_i |A_i P_Y A_i^T|^{-1/2} \exp\left\{ -\frac{1}{2} \tilde{w}_i^T (A_i P_Y A_i^T)^{-1} \tilde{w}_i \right\}$$

$$\Gamma = \sum_{i=1}^N \beta_i$$
\[
\gamma_i = \frac{\beta_i}{\Gamma} \tag{119}
\]
then we can rewrite Eq. (115) as follows:
\[
f_{X|W}(x|w) = (2\pi)^{-n/2} \frac{1}{\Gamma} \sum_{i=1}^{N} \gamma_i \exp \{ -\frac{1}{2} (x - \frac{1}{2} x^T S_{X} x) \}
\]
\[
= (2\pi)^{-n/2} \frac{1}{\Gamma} \sum_{i=1}^{N} \gamma_i \frac{\beta_i}{\Gamma} \tag{120}
\]
\[
= (2\pi)^{-n/2} \sum_{i=1}^{N} \gamma_i \frac{\beta_i}{\Gamma} \tag{121}
\]
\[
= (2\pi)^{-n/2} \sum_{i=1}^{N} \gamma_i \delta_i \exp \{ -\frac{1}{2} x^T S_{X} x + x^T S_{XW} \hat{w}_i + \hat{w}_i^T S_{WX} x \}
\]
\[
+ \hat{w}_i^T [S_{WW} - (A_i P_Y A_i^T)^{-1}] \hat{w}_i \} \tag{123}
\]
where \( \delta_i = |\tilde{A}_i P_Z \tilde{A}_i^T|^{-1/2} |A_i P_Y A_i^T|^{1/2} \) and
\[
S = \begin{bmatrix}
S_X & S_{XW} \\
S_{XW} & S_W
\end{bmatrix} = (\tilde{A}_i P_Z \tilde{A}_i^T)^{-1} \tag{125}
\]
expanding \( S(\tilde{A}_i P_Z \tilde{A}_i^T) = I \) yields
\[
S_X P_X + S_{XW} A_i P_Y X = I \tag{126}
\]
\[
S_X P_X + S_{XW} A_i P_Y A_i^T = 0 \tag{127}
\]
\[
S_{WX} P_X + S_{WW} A_i P_Y X = 0 \tag{128}
\]
\[
S_{WX} P_X + S_{WW} A_i P_Y A_i^T = I \tag{129}
\]
Completing the square in the exponent yields
\[
f_{X|W}(x|w) = (2\pi)^{-n/2} \sum_{i=1}^{N} \gamma_i \delta_i \exp \{ -\frac{1}{2} (x + S_{X}^{-1} S_{XW} \hat{w}_i) \}
\]
\[
+ \hat{w}_i^T [S_{WW} - S_{WX} S_{X}^{-1} S_{XW} - (A_i P_Y A_i^T)^{-1}] \hat{w}_i \} \tag{130}
\]
\[
S_{X}^{-1} = P_X - P_{XY} A_i^T (A_i P_Y A_i^T)^{-1} A_i P_{YX} \tag{131}
\]
\[
S_{X}^{-1} S_{XW} = -P_{XY} A_i^T (A_i P_Y A_i^T)^{-1} \tag{132}
\]
\[
0 = S_{WW} - S_{WX} S_{X}^{-1} S_{XW} - (A_i P_Y A_i^T)^{-1} \tag{133}
\]
Substituting these into Eq. (130) yields
\[
f_{X|W}(x|w) = (2\pi)^{-n/2} \sum_{i=1}^{N} \gamma_i \delta_i \exp \{ -\frac{1}{2} (x - B_i w) \}
\]
where \( B_i = -P_{XY} A_i^T (A_i P_Y A_i^T)^{-1} \) and \( P_i = P_X - P_{XY} A_i^T (A_i P_Y A_i^T)^{-1} A_i P_{YX} \). According to block matrix determinant algebra, \( \delta_i \) turns up to be \( |P_i|^{-1/2} \), and according to the definition
of $\gamma_i$, we have that $0 \leq \gamma_i \leq 1$ and $\sum_{i=1}^{N} \gamma_i = 1$. This form coincides with the form of Eq. (83), thus the optimal estimation of $X$ by given $W$ is as follows:

$$\hat{x} \triangleq E\{X|W = w\}$$

$$= \int x f_{X|W}(x|w) \, dx$$

$$= \sum_{i=1}^{N} \gamma_i B_i \bar{w}_i \quad \text{(134)}$$

and the covariance of the estimation error is

$$P \triangleq E\{\|X - \hat{x}\|^2|W = w\}$$

$$= \sum_{i=1}^{N} \gamma_i^2 P_i \quad \text{(135)}$$

References

