Estimation from Lossy Sensor Data:
Jump Linear Modeling and Kalman Filtering

Alyson K. Fletcher
alyson@eecs.berkeley.edu
Univ. of California, Berkeley
Berkeley, CA 94720, USA

Sundeep Rangan
s.rangan@flarion.com
Flarion Technologies
Bedminster, NJ 07921, USA

Vivek K Goyal
vgoyal@mit.edu
Massachusetts Inst. of Tech.
Cambridge, MA 02139, USA

ABSTRACT

Due to constraints in cost, power, and communication, losses often arise in large sensor networks. The sensor can be modeled as an output of a linear stochastic system with random losses of the sensor output samples. This paper considers the general problem of state estimation for jump linear systems where the discrete transitions are modeled as a Markov chain. Among other applications, this rich model can be used to analyze sensor networks. The sensor loss events are then modeled as Markov processes. Under the jump linear system model, many types of underlying losses can be easily considered, and the optimal estimator to be performed at the receiver in the presence of missing sensor data samples is given by a standard time-varying Kalman filter.

We show that the asymptotic average estimation error variance converges and is given by a Linear Matrix Inequality, which can be easily solved. Under this framework, any arbitrary Markov loss process can be modeled, and its average asymptotic error variance can be directly computed. We include a few illustrative examples including fixed-length burst errors, a two-state model, and partial losses due to multiple SNR states. Our analysis encompasses modeling discrete changes not only in the received data as stated above, but also in the underlying system. In the context of the lossy sensor model, the former allows for variation in sensor positioning, power control, and loss of data communications; the latter could allow for discrete changes in the dynamics of the variable monitored by the sensor. This freedom in modeling yields a tool that is potentially valuable in various scenarios in which entities that share information are subjected to challenging and time-varying network conditions.

Categories and Subject Descriptors

H.1.1 [**Systems and Information Theory**]: General systems theory

General Terms

Algorithms

Keywords

loss mitigation, minimum mean-squared error estimation, signal recovery, state-space modeling, tracking

1. INTRODUCTION

The possibility of deploying small, cheap, networked sensors in large numbers is a relatively recent concept and yet has inspired an extraordinary amount of research [24]. One reason for this is that the range of research contributions that pertain in some way to sensor networks is vast. For example, it includes a variety of issues related to building sensors (miniaturization, ruggedization, improving measurement and communication abilities, lowering power consumption, etc.), to the communication between sensors and other network nodes (ad hoc networking, fault tolerance, scheduling, multi-terminal information theory, etc.), and to using data collected by a sensor network (estimation, inference, classification, hypothesis testing, etc.).

This paper addresses centralized estimation of a time-varying real vector from the measurements made at sensors distributed across a network, where those measurements are subject to communication loss and to degradation from sensor noise and quantization. In particular, we provide a computational framework for estimating the asymptotic estimation error variance without resorting to stochastic simulation. A straightforward application of this framework provides a sensor network system designer with an efficient means to see the effects of communication reliability and sensor resolution on estimation quality. It may also be useful in optimizing sensor density and placement. Of course, sensor networks are not the only place where such estimation problems arise, but we emphasize these applications for these Proceedings. Another application is discussed briefly in our concluding comments.

1.1 Estimation with Physical Dynamics

In our sensor network model, "messages" are sent from sensors to data collection nodes. Each data collection node is attempting to estimate a real vector quantity from the messages. The data collection nodes operate independently, so we henceforth consider just one node. Our interest is in relating the *physical dynamics* of the measured quantity and the *communication reliability* of the network to the quality of estimate that is obtained, with great generality. We

propose to use a fairly rich set of models for the physical process—jump linear systems. Jump linear systems are described formally in Section 5; informally, a jump linear system is a linear, time-varying system in which the dynamics of the system "jump" between discrete states, within any one discrete state the system is linear and time invariant, and the transition between states satisfy the Markov property. For jump linear systems we study the mean-squared error (MSE) of optimal estimators when the measurements are subject to various types of degradations.

A contrasting approach, which is more information-theoretic in style, is to design the transmission strategy jointly for all the sensors to optimize the estimation [13].[1] While this alternative is promising, we address instead the traditional system architectures with separation between an application/data layer vs. a channel/network layer. Thus far, nonstationary signal models have not been incorporated into the communication-centric approach; conversely, our approach applies for autoregressive models with Markov regimes.

The communication between sensors and data collection nodes in a sensor network is usually erratic. This may be due to time-variation in sensor positions or the communication media, or allocating power and bandwidth sufficient for transparent, lossless communication may be inefficient. It may also be caused by sensors turning themselves off to conserve energy or by device failures. Sensor network communication models—which may include, for example, measurement losses with various correlations—prompt the development of our new estimation results. The theoretical context for this work comes from the theory of stochastic linear systems [16]. This field provides a large repository of results to build upon; however, the most attention has been focused on *control* rather than *estimation* and *continuous time* rather than *discrete time*—the duals of what we are most interested in. In particular, an attractive feature of our model is that the dual problem of control of jump linear systems has been extensively analyzed. Our approach also makes extensive use of a convex programming method based on linear matrix inequalities (LMIs), a technique widely used in control [2].

1.2 Overview and Paper Structure

The initial contribution of this paper is to demonstrate that jump linear systems are a good modeling framework for estimation problems that arise with measurements subject to various sorts of Markov degradation processes. Then, the main contribution is to show how to relate the state estimation error covariance and output estimation error variance of these systems to a constrained optimization with LMIs. These optimization problems can be numerically solved very efficiently and in some cases solved symbolically. One application of this work could be to optimize source and channel coding rates in a sensor network according to an error variance criterion.

This paper is structured as follows: First, previous work is outlined in Section 2. The state-space signal model and overall system model are given in Section 3, followed by a review of optimal estimation in Section 4. The Markov loss model described in Section 5 is central to the development

[1]While the transmissions of sensors are designed for a global criterion, communication between sensors may or may not be present.

because it turns the overall system into a Markov jump linear system. Several examples are given. Section 6 then gives the main results of the paper, which are characterizations of the asymptotic state estimation error covariance and output estimation error variance in terms of LMIs. Numerical examples are given in Section 7.

2. PREVIOUS WORK

Jump linear systems have been studied extensively in the field of systems and control and a bit less so in signal processing. A continuous-time quadratic jump linear control problem appeared as early as the 1960s in the work of Krasovskii and Lidskii [14]. Sworder [22] and Wonham [23] provided early solutions to the problem using optimal control and dynamic programming methods. Since then, numerous extensions have been considered. In particular, the discrete-time control problem is examined in several papers including [3, 6, 5, 7, 18]. The solution presented in [3, 5, 7] is based on a coupled versions of the standard Riccati equation arising in optimal control of LTI systems (see, for example, [8]). LMI solutions to the coupled Riccati equations were developed in [6] and [18].

In this paper, we consider the jump linear state estimation problem which, mathematically, is the dual of the state-feedback control problem. We exploit this duality to derive the main result of this paper, Theorem 1, which presents an LMI similar to that in [6].

Jump linear system state estimation is also considered in [1, 4, 9, 10, 11, 15, 17]. However, these papers consider the more difficult problem of estimating both the linear system state *and* the discrete Markov state from a noisy output. In this paper, the Markov state is assumed to be known, and we compute the average estimation error of the Kalman filter. While the problem we look at is easier, the focus is not on presenting estimation methods but rather on being able to efficiently compute or approximate the *performance* of optimal estimators.

As for sensor modeling, the use of Kalman filtering to handle sensors has been proposed for vehicular sensors in [19] and sensor network problems in [20]. A jump linear system model with Markov sensor errors is presented in [21]. The paper [20] provides an LMI analysis, but only for the case of independent and identically distributed (i.i.d.) sensor erasures. This paper can be seen as an extension of this LMI analysis to more general signal models and general Markov error models.

3. STATE-SPACE SENSOR MODELING: STATIONARY SIGNAL CASE

We consider the lossy sensor system in Fig. 1(a). The signal detected by the sensor is denoted $z(k)$. All signals are in discrete time. The sensor produces an output $y_0(k)$, which may differ from $z(k)$ due to sensor noise and other sensor imperfections. The sensor output samples are transmitted over a lossy channel, which can result in sample erasures. A receiver receives the channels $y(k)$ and must estimate the original signal $z(k)$ from whatever it receives. The estimator must account for the sensor imperfections and any losses in the channel. The estimator output is denoted $\hat{z}(k)$.

The signal of interest, $z(k)$, is modeled as a correlated random process. The process correlations are assumed to be known at the receiver and can be exploited to estimate
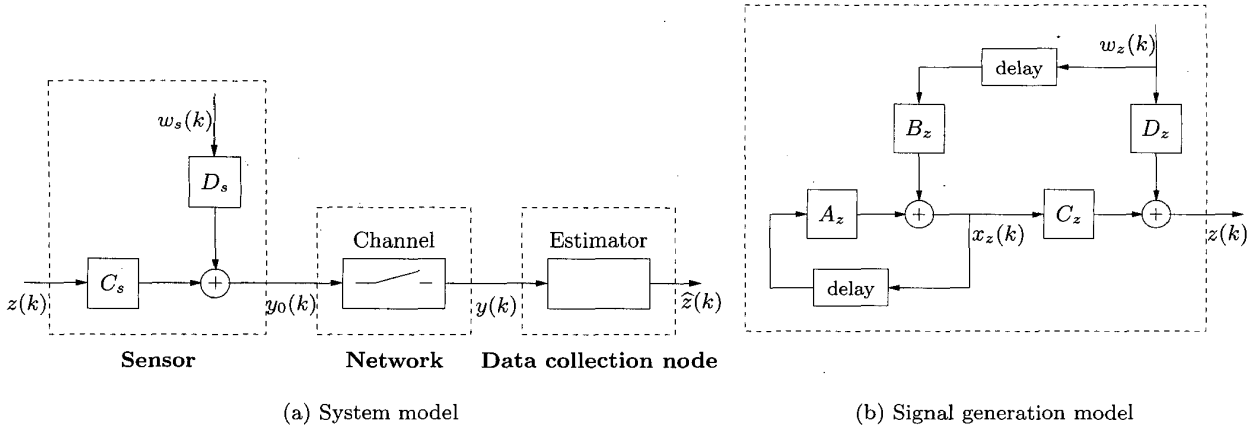
(a) System model

(b) Signal generation model

Figure 1: (a) Overall system model. Desired signal $z(k)$ is observed by a sensor that generates $y_0(k)$. The network is abstracted as an erasure channel that passes $y(k)$ to the data collection node. The result of estimation is denoted $\hat{z}(k)$. (b) Signal generation model whereby $z(k)$ is a filtered version of white noise $w_z(k)$. The state-space model (A_z, B_z, C_z, D_z) is also denoted as H_z.

$z(k)$ from the lossy sensor samples. Specifically, we use the standard and general model whereby $z(k)$ is assumed to be filtered white noise. This abstract model is shown in Fig. 1(b), where $w_z(k)$ is unit variance noise and H_z is a linear time-invariant filter. Any finite-order wide-sense stationary process can be represented in such a manner with an appropriate choice of H_z; consequently, this model is widely-used. Typically, $z(k)$ is assumed to be slowly-varying, and H_z is selected to be low-pass with a cutoff frequency with appropriate bandwidth. An example is illustrated in the numerical examples at the end of the paper.

The filter H_z can itself be described in numerous ways. For the purposes of this paper, it will be most convenient to represent H_z in state space form as in Fig. 1(b):

$$H_z : \quad \begin{matrix} x_z(k+1) & = & A_z x_z(k) + B_z w_z(k) \\ z(k) & = & C_z x_z(k) + D_z w_z(k) \end{matrix} \quad (1)$$

where $x_z(k)$ is the system state and the state-space matrices (A_z, B_z, C_z, D_z) describe the filter H_z.

Turning to the sensor, we model the sensor as a linear gain with additive noise. Specifically, we assume that

$$y_0(k) = C_s z(k) + D_s w_s(k), \quad (2)$$

where $w_s(k)$ is unit-variance white noise, C_s is a linear gain term, and D_s is a noise scaling term depending on the sensor noise variance. The resulting sensor noise variance is $D_s D_s'$. The sensor noise $w_s(k)$ is assumed to account for sensor errors and any finite resolution (quantization) and compression effects. The sensor noise is shown in the abstract model in Fig. 1(b) as a random input to the sensor.

Finally, for the channel, channel losses are modeled as sample erasures. That is, each sample $y_0(k)$ is either received perfectly or completely lost and unavailable to the receiver:

$$y(k) = \begin{cases} y_0(k) & \text{if sample } k \text{ is not lost;} \\ 0 & \text{if sample } k \text{ is lost.} \end{cases} \quad (3)$$

This erasure model is well-suited for digital communication of the sensor data where the samples $y_0(k)$ are digitally en-

coded and transmitted, and channel losses result in complete losses of the digital data for one or more samples.

Example: A simple example of this framework could have

$$\begin{aligned} A_z &= \alpha \in (0,1) \\ B_z &= (1-\alpha^2)^{1/2}\sigma_z \in \mathbb{R} \\ C_z &= 1 \\ D_z &= 0 \\ C_s &= 1 \\ D_s &= \sigma_n \in \mathbb{R}, \end{aligned}$$

which represents a first-order autoregressive signal $z(k)$ with total power $\mathbf{E}|z(k)|^2 = \sigma_z^2$. The parameter α determines the correlation of the process $z(k)$: A process with $\alpha = 0.95$ would be slowly varying and a process with $\alpha = 0.6$ would vary more quickly. The parameter σ_n^2 is the sensor noise power, so the sensor SNR is σ_z^2/σ_n^2.

Note: In Section 5, we extend the signal model by replacing the linear, time-invariant system H_z with a jump linear system, i.e., a system whose state includes an M-state Markov chain and a potentially different (A_z, B_z, C_z, D_z) for each state. We avoid that additional complication at present to simplify the notation in Section 4.

4. KALMAN FILTERING AND ESTIMATION REVIEW

In the previous section, the receiver must estimate the sensor input $z(k)$ from the lossy channel output $y(k)$. The main motivation for the linear state-space modeling is that the optimal estimator is given by a standard Kalman filter. The Kalman filter is classical and described in numerous texts such as [16].

The Kalman filter provides a recursive method for computing the minimum mean square error (MMSE) estimates of the state $x_z(k)$ and output $z(k)$ from the output $y(k)$.

The MMSE estimates are the conditional expectations

$$\begin{aligned}
\widehat{x}_{opt}(k\,|\,k-1) &= \mathbf{E}(x_z(k)\,|\,y_0^{k-1}), \\
\widehat{z}_{opt}(k\,|\,k-1) &= \mathbf{E}(z(k)\,|\,y_0^{k-1}),
\end{aligned} \tag{4}$$

where y_0^{k-1} denotes the subsequence $y(0)$, $y(1)$, ..., $y(k-1)$. That is, the estimates $\widehat{x}_{opt}(k\,|\,k-1)$ and $\widehat{z}_{opt}(k\,|\,k-1)$ are the MMSE estimates of the state $x_z(k)$ and sensor input $z(k)$ given the output samples $y(j)$ up to time $k-1$. We use the subscript "opt" to indicate that these estimates are, by definition, the optimal estimates in the MMSE sense. Below, we will compare these estimates with certain suboptimal estimators. Let $P_{opt}(k\,|\,k-1)$ denote the state estimation error covariance and $J_{opt}(k\,|\,k-1)$ the mean square output estimation error:

$$\begin{aligned}
P_{opt}(k\,|\,k-1) &= \mathbf{E}\left[(x_z(k) - \widehat{x}_{opt}(k))(x_z(k) - \widehat{x}_{opt}(k))'\right] \\
J_{opt}(k\,|\,k-1) &= \mathbf{E}\left[\|z(k) - \widehat{z}_{opt}(k)\|^2\right].
\end{aligned} \tag{5}$$

To employ the Kalman filter in the lossy sensor problem, we first need to combine the signal model H_z and the sensor into a single state-space system. To this end, we define the joint noise vector

$$w(k) = [w_z(k)' \;\; w_s(k)']'.$$

Now, combining (1) with (2), we can write the signal $z(k)$ and sensor output $y_0(k)$ as outputs of a single state-space system:

$$\begin{aligned}
x(k+1) &= Ax(k) + Bw(k), \\
z(k) &= C_1 x(k) + D_1 w(k), \\
y_0(k) &= C_2 x(k) + D_2 w(k),
\end{aligned} \tag{6}$$

where

$$\begin{aligned}
A &= A_z \\
B &= [B_z \quad 0] \\
C_1 &= C_z \\
D_1 &= [D_z \quad 0] \\
C_2 &= C_s C_z \\
D_2 &= [C_s D_z \quad D_s].
\end{aligned}$$

Using this state space model, we can now state the standard Kalman filter equations. The reader is referred to [16] for derivations.

The Kalman equations are recursive in that the state and error covariance estimates at each sample k depend on the estimates at the previous sample $k-1$. In the sensor problem, there are two cases for the updates, corresponding to whether the sample $y_0(k)$ is lost or not. If the sample is not lost, the state and state error covariance are updated with the equations

$$\begin{aligned}
\widehat{x}_{opt}(k+1\,|\,k) &= A\widehat{x}_{opt}(k\,|\,k-1) \\
&\quad + L(k)(y(k) - C_2\widehat{x}_{opt}(k\,|\,k-1)), \\
P_{opt}(k+1\,|\,k) &= AP_{opt}(k\,|\,k-1)A' + BB' \\
&\quad - L(k)G(k)L(k),
\end{aligned} \tag{7}$$

where

$$\begin{aligned}
G(k) &= C_2 P_{opt}(k\,|\,k-1)C_2' + D_2 D_2', \\
L(k) &= (AP_{opt}(k\,|\,k-1)C_2' + BD_2')G(k)^{-1}.
\end{aligned}$$

Alternatively, if the sample k is lost,

$$\begin{aligned}
\widehat{x}_{opt}(k+1\,|\,k) &= A\widehat{x}_{opt}(k\,|\,k-1) \\
P_{opt}(k+1\,|\,k) &= AP_{opt}(k\,|\,k-1)A' + BB'.
\end{aligned} \tag{8}$$

In both cases, the MMSE estimate for $z(k)$ and the output error variance are given by

$$\begin{aligned}
\widehat{z}_{opt}(k\,|\,k-1) &= C_1\widehat{x}_{opt}(k\,|\,k-1), \\
J_{opt}(k\,|\,k-1) &= \mathbf{Tr}\left[C_1 P_{opt}(k\,|\,k-1)C_1' + D_1 D_1'\right].
\end{aligned} \tag{9}$$

5. MARKOV LOSS MODEL

The primary application of the framework of this paper is to quantify the effect of channel losses on the sensor estimation error. To do this, we need a statistical model for the erasures on the channel.

We model the erasures with a *hidden Markov model*. That is, we assume that there is some Markov chain

$$\theta(k) \in \{1, 2, \ldots, M\} \qquad \text{for } k \in \mathbb{Z}$$

such that the sample $y_0(k)$ is lost if and only if $\theta(k) \in I_{\text{loss}}$. I_{loss} is some subset of the discrete states $\{1, 2, \ldots, M\}$. We will let p_{ij} denote the transition probabilities

$$p_{ij} = \mathbf{Pr}(\theta(k+1) = j\,|\,\theta(k) = i). \tag{10}$$

We will assume that the chain is aperiodic and irreducible, so that it admits a unique stationary distribution

$$q_i = \mathbf{Pr}(\theta(k) = i), \tag{11}$$

which is given by the unique solution to the equations

$$q_j = \sum_{i=1}^{M} q_i p_{ij}, \qquad j = 1, 2, \ldots, M.$$

The hidden Markov model is extremely general and can incorporate a wide range of loss processes. The modeling is best illustrated with some simple examples:

- *Independent losses:* In this case, we use $M = 2$ states. One state, say $\theta(k) = 1$, corresponds to a loss of the sample $y_0(k)$, and the other state $\theta(k) = 2$ corresponds to a non-loss. If $\theta(k)$ is i.i.d., a loss probability of λ can be represented with the transition probabilities

$$p_{1,1} = p_{2,1} = 1 - p_{1,2} = 1 - p_{2,2} = \lambda.$$

The stationary distribution of the Markov chain is

$$q_1 = 1 - q_2 = \lambda. \tag{12}$$

- *Gilbert-Elliot losses:* This loss model is identical to the model above, except that the transition probabilities are

$$\begin{aligned}
p_{1,1} &= 1 - p_{1,2} = \lambda_1 \\
p_{2,1} &= 1 - p_{2,2} = \lambda_2
\end{aligned}$$

for some probabilities λ_1 and λ_2. This provides a simple model for correlated and bursty losses. The parameter λ_2 represents the probability of going from the non-lossy to the lossy state. Once in the lossy state, the system will remain in the lossy state an exponentially distributed time with mean $1/(1-\lambda_1)$. The stationary distribution of the Markov chain is

$$q_1 = 1 - q_2 = \frac{\lambda_2}{1 - \lambda_1 + \lambda_2}. \tag{13}$$

The value of q_1 in (13) is the overall loss probability; it reduces to λ for $\lambda_1 = \lambda_2 = \lambda$ as expected.

- *Fixed-length burst losses:* As another example of burst losses, suppose that when losses occur, a fixed number, L, consecutive samples of $y_0(k)$ are lost. This differs from the Gilbert-Elliot model where the number of samples lost at a time are random. Also, suppose that, on any given sample, an L-length burst loss occurs with some probability $\lambda > 0$. To model this type of loss, we use an $L + 1$ state Markov chain, so that $\theta(k) \in \{1, \ldots, L + 1\}$. Again, we use one state, say $\theta(k) = L + 1$ to correspond to the state with no loss. In the other L states, $\theta(k) = i$ indicates that the current sample is lost along with the i previous samples in the L-length burst loss. Using these states, the transition probabilities for this chain are of the form

$$
\begin{aligned}
p_{i,1} &= \lambda && \text{for } i = 1, \ldots, L + 1, \\
p_{i,i+1} &= 1 - \lambda && \text{for } i = 1, \ldots, L, \\
p_{L+1,L+1} &= 1 - \lambda,
\end{aligned}
$$

with all other transition probabilities equal to zero.

The above are just a few examples. As described in [16], by appropriately discretizing the distribution, virtually any distribution on the number of losses can be modeled with enough states.

The Markov loss process—and indeed many other modeling possibilities—are captured by replacing the state space system (6) with the following *jump linear system*:

$$
\begin{aligned}
x(k+1) &= A_{\theta(k)}x(k) + B_{\theta(k)}w(k), \\
z(k) &= C_{\theta(k),1}x(k) + D_{\theta(k),1}w(k), \\
y(k) &= C_{\theta(k),2}x(k) + D_{\theta(k),2}w(k).
\end{aligned} \tag{14}
$$

The particular case we have focused on has

$$
\begin{aligned}
A_{\theta(k)} &= A && \text{for all } \theta(k), \\
B_{\theta(k)} &= B && \text{for all } \theta(k), \\
C_{\theta(k),1} &= C_1 && \text{for all } \theta(k), \\
D_{\theta(k),1} &= D_1 && \text{for all } \theta(k), \\
C_{\theta(k),2} &= \begin{cases} C_2 & \text{for } \theta(k) \notin I_{\text{loss}}; \\ 0 & \text{for } \theta(k) \in I_{\text{loss}}, \end{cases} \\
D_{\theta(k),2} &= \begin{cases} D_2 & \text{for } \theta(k) \notin I_{\text{loss}}; \\ 0 & \text{for } \theta(k) \in I_{\text{loss}}. \end{cases}
\end{aligned}
$$

Having a plurality of modes for the physical dynamics would give a new, larger discrete state space and, potentially, dependence on $\theta(k)$ for $A_{\theta(k)}$, $B_{\theta(k)}$, $C_{\theta(k),1}$, and $D_{\theta(k),1}$. A set of sensor modes (e.g., different resolutions or quantizations) would also increase the state space and multiply the possibilities for $C_{\theta(k),2}$ and $D_{\theta(k),2}$.

6. LMI ASYMPTOTIC AVERAGE ERROR ANALYSIS

6.1 Asymptotic Average Error

In Section 4, the Kalman filter updates depend on which particular samples are lost, and the sample losses are in turn determined by the Markov parameter $\theta(k)$. Consequently, since $\theta(k)$ is a random process, the state estimation error covariance $P_{opt}(k \mid k-1)$ and output estimation error variance $J(k \mid k-1)$ will also be random processes varying with $\theta(k)$. This property will be true in a general jump linear system as well: the state and output estimation error variance

will be time-varying quantities depending on the particular realization of the Markov parameter $\theta(k)$.

The goal of this paper is to estimate the *asymptotic average* estimation error. For the general jump linear system (14), consider the state and output estimates

$$
\begin{aligned}
\widehat{x}_{opt}(k \mid k-1) &= \mathbf{E}(x_z(k) \mid y_0^{k-1}, \theta_0^k), \\
\widehat{z}_{opt}(k \mid k-1) &= \mathbf{E}(z(k) \mid y_0^{k-1}, \theta_0^k),
\end{aligned} \tag{15}
$$

which are identical to the estimates in (4), except that we have explicitly indicated the dependence of the estimates on $\theta(k)$. As in the particular case of the sensor model example, the estimates $\widehat{x}_{opt}(k \mid k-1)$ and $\widehat{z}_{opt}(k \mid kmm)$ can be computed with a time-varying Kalman filter. Let $P_{opt}(k \mid k-1)$ denote the corresponding state estimation error covariance and $J_{opt}(k \mid k-1)$ the mean square output estimation error:

$$
\begin{aligned}
&P_{opt}(k \mid k-1) \\
&= \mathbf{E}\left[(x_z(k) - \widehat{x}_{opt}(k))(x_z(k) - \widehat{x}_{opt}(k))' \mid \theta_0^k\right], \\
&J_{opt}(k \mid k-1) = \mathbf{E}\left[\|z(k) - \widehat{z}_{opt}(k)\|^2 \mid \theta_0^k\right].
\end{aligned}
$$

We define the asymptotic average state error covariance and mean square output error as

$$
\begin{aligned}
\overline{P}_{opt} &= \lim_{k \to \infty} \mathbf{E}\left[P_{opt}(k \mid k-1)\right], \\
\overline{J}_{opt} &= \lim_{k \to \infty} \mathbf{E}\left[J_{opt}(k \mid k-1)\right],
\end{aligned} \tag{16}
$$

where the expectations are over the possible random trajectories $\theta(k)$. Since $P_{opt}(k \mid k-1)$ is the state estimate covariance conditional on $\theta(j)$ up to time $j = k$, it follows that

$$
\overline{P}_{opt} = \lim_{k \to \infty} \mathbf{E}(x(k) - \widehat{x}_{opt}(k \mid k-1))(x(k) - \widehat{x}_{opt}(k \mid k-1))'.
$$

That is, $\overline{P}_{opt}$ is the asymptotic average minimal state variance, averaging over both the noise $w(k)$ and discrete state sequence $\theta(k)$. Similarly, we define

$$
\overline{J}_{opt} = \lim_{k \to \infty} \mathbf{E}\|z(k) - \widehat{z}_{opt}(k \mid k-1)\|^2,
$$

so $\overline{J}_{opt}$ represents the average mean square output error.

The limit as $k \to \infty$ in (16) is to eliminate transient effects of the initial conditions in either the Markov chain or linear state-space system.

6.2 Suboptimal Estimator and LMI Analysis

In general, it is difficult to compute the asymptotic average errors, $\overline{P}_{opt}$ and $\overline{J}_{opt}$, exactly for the Kalman filter. Instead, we consider a slightly suboptimal estimator for which we can compute the error easily through an LMI. The error of the suboptimal estimator will serve as an upper bound on the minimum error achieved by the Kalman filter.

To this end, consider a state and output estimator of the form

$$
\begin{aligned}
\widehat{x}(k+1) &= A_i\widehat{x}(k) + L_i(y(k) - C_{i2}\widehat{x}(k)), \\
\widehat{z}(k) &= C_{i1}\widehat{x}(k),
\end{aligned} \tag{17}
$$

where $i = \theta(k)$. The matrices $L_1, L_2, \ldots, L_M$ will be called the estimator *gain matrices* and represent adjustable parameters. The estimator (17) is identical in form to the Kalman estimator in (7) except for these gain matrices. In the Kalman filter, the gain matrix $L(k)$ in (7) is, in effect, a function of all the values of $\theta(j)$ up to time $j = k$. In the estimator (17), $L(k)$ is restricted to be a function only of the

current value of $\theta(k)$. In this way, the class of estimators defined by (17) may be slightly suboptimal since it may not necessarily include the optimal Kalman estimator.

The advantage of the estimator (17) is that it can be analyzed and optimized easily using LMIs. We first analyze the performance of the estimator for a given set of gain matrices, $L_1, L_2, \ldots, L_M$. The analysis in this case is standard: Define the state and output estimation error sequences

$$e_x(k) = x(k) - \widehat{x}(k) \quad \text{and} \quad e_z(k) = z(k) - \widehat{z}(k).$$

Using this notation, the state and output estimation errors for the system (17) are given by

$$
\begin{aligned}
P(k) &= \mathbf{E}\left[(x(k) - \widehat{x}(k))(x(k) - \widehat{x}(k))'\right] \\
&= \mathbf{E}\left[e_x(k)e_x(k)'\right], \\
J(k) &= \mathbf{E}\left[\|z(k) - \widehat{z}(k)\|^2\right] \\
&= \mathbf{E}\left[\|e_z(k)\|^2\right],
\end{aligned}
$$

and the corresponding asymptotic average values are

$$\overline{P} = \lim_{k \to \infty} P(k) \quad \text{and} \quad \overline{J} = \lim_{k \to \infty} J(k). \tag{18}$$

Also, combining (17) and (14), we see that

$$
\begin{aligned}
e_x(k+1) &= A_{cl,i}e_x(k) + B_{cl,i}w(k) \\
e_z(k) &= C_{i1}e_x(k) + D_{i1}w(k)
\end{aligned} \tag{19}
$$

where $i = \theta(k)$ and the subscript "cl" is used to denote the "close-loop" matrices

$$A_{cl,i} = A_i - L_iC_{i2} \quad \text{and} \quad B_{cl,i} = B_i - L_iD_{i2}.$$

Equations (19) show that the state and output error sequences $e_x(k)$ and $e_z(k)$ are themselves the state and output of a jump linear state-space system. The asymptotic errors in (18) are now given by the following standard LMI.

LEMMA 1 ([6]). *Consider the jump linear system (14) and corresponding estimator (17) defined for a set of gain matrices $L_1, L_2, \ldots, L_M$. Then, the asymptotic mean-square output estimation error $\overline{J}$, defined in (18), is given by*

$$\overline{J} = \min \sum_{i=1}^{M} q_i \mathbf{Tr}(C_{i1}P_iC_{i1}' + D_{i1}D_{i1}')$$

where the minimization is over matrices

$$P_i \geq 0, \qquad i = 1, 2, \ldots, M$$

satisfying the coupled Lyapunov inequalities

$$P_i \geq A_{cl,i}\overline{P}_iA_{cl,i}' + B_{cl,i}B_{cl,i}', \tag{20}$$

where

$$\overline{P}_i = \sum_{j=1}^{M} p_{ij}P_j.$$

(Recall that the q_is are the stationary distribution of the Markov chain $\theta(k)$.) Also, if P_is are the minimizing matrices, the asymptotic state error covariance $\overline{P}$, defined in (18), is given by

$$\overline{P} = \sum_{i=1}^{M} q_i P_i.$$

If there is no feasible set of P_is satisfying the coupled Lyapunov inequalities (20), then the limits (18) do not converge.

The minimization in Lemma 1 can be solved easily as an LMI in the variables P_i, thereby providing a way of computing the asymptotic average errors $\overline{P}$ and $\overline{J}$.

The following lemma, which is also standard, provides an alternative LMI for computing $\overline{J}$. Although the state error variance $\overline{P}$ cannot be computed with this method, the method is useful in the next subsection.

LEMMA 2 ([6]). *Consider the jump linear system (14) and corresponding estimator (17) defined for a set of gain matrices $L_1, L_2, \ldots, L_M$. Then, the asymptotic mean-square output estimation error $\overline{J}$, defined in (18), is given by*

$$\overline{J} = \min \sum_{i=1}^{M} q_i \mathbf{Tr}(B_{cl,i1}'Q_iB_{cl,i1} + D_{i1}'D_{i1})$$

where the minimization is over matrices

$$Q_i \geq 0, \qquad i = 1, 2, \ldots, M$$

satisfying the coupled Lyapunov inequalities

$$Q_i \geq A_{cl,i}'\overline{Q}_iA_{cl,i} + C_{i1}'C_{i1}, \tag{21}$$

where

$$\overline{Q}_i = \sum_{j=1}^{M} p_{ij}Q_j.$$

If there is no feasible set of Q_is satisfying the coupled Lyapunov inequalities (21), then the limits (18) do not converge.

6.3 Estimation Optimization and the LMI Upper Bound

The previous subsections show how to compute the estimation error for the estimator (17) for a given set of gain matrices $L_1, L_2, \ldots, L_M$. In this subsection, we show how the optimal gain matrices can be found also using an LMI.

The optimization is based on finding matrices $W_i > 0$ satisfying the following *coupled Riccati inequalities*:

$$
\begin{aligned}
A_i'\overline{W}_{11}A_i &+ C_{i2}'\overline{W}_{i2}'A_i + A_i'\overline{W}_{i2}C_{i2} \\
&+ C_{i2}'\overline{W}_{i3}C_{i2} + C_{i1}'C_{i1} - W_{i1} \leq 0,
\end{aligned} \tag{22}
$$

where W_i is partitioned as

$$W_i = \begin{pmatrix} W_{i1} & W_{i2} \\ W_{i2}' & W_{i3} \end{pmatrix}$$

and

$$\overline{W}_i = \sum_{j=1}^{M} p_{ij}W_j.$$

We have the following theorem.

THEOREM 1. *Consider the jump linear system (14). Suppose that there exist matrices $W_i > 0$ satisfying the coupled Riccati LMIs (22) and*

$$L_i = -W_{i1}^{-1}W_{i2}.$$

Then, for the estimator (17), the asymptotic mean square output error $\overline{J}$, defined in (18), is bounded by

$$\overline{J} \leq \sum_{i=1}^{M} q_i \mathbf{Tr}([B_i' \ D_{i2}'] W_i [B_i' \ D_{i2}']') + D_{i1}D_{i1}'. \tag{23}$$

Conversely, if there exist matrices L_i resulting in an asymptotic mean square output error $\overline{J}$, there must exist matrices

$W_i \geq 0$ *satisfying the coupled Riccati LMIs (22) and the upper bound (23).*

Theorem 1 shows that the optimal gain matrices L_i for the estimator (17) can be found by minimizing the right hand side of (23) over the matrix variables W_i subject to Riccati LMIs (22). This minimization is an LMI.

The theorem can also be used to bound the performance of the Kalman filter. Recall that the Kalman filter is the optimal estimator in MMSE sense. Therefore, for any gain matrices L_i, the mean square output error $\overline{J}$ of the estimator (17) cannot be lower than the corresponding error $\overline{J}_{opt}$ of the Kalman filter. That is, for all gain matrices L_i,

$$\overline{J}_{opt} \leq \overline{J}.$$

Combining this fact with Theorem 1, we have the following bound on the average error of the Kalman filter.

COROLLARY 1. *Consider the jump linear system (14) and corresponding Kalman filter. Then, if there exists matrices $W_i > 0$ satisfying the coupled Riccati LMIs (22), the Kalman filter asymptotic mean square output error $\overline{J}_{opt}$ in (16) is bounded by*

$$\overline{J}_{opt} \leq \sum_{i=1}^{M} q_i \mathrm{Tr}([B_i' \ D_{i2}'] W_i [B_i' \ D_{i2}']') + D_{i1} D_{i1}'.$$

7. NUMERICAL EXAMPLES

To demonstrate the types of calculations that can be made with this framework, we provide a simple example generated with the Gilbert-Elliot loss model.

Following the notation in Section 3, we take the signal $z(k)$ to be a low-pass Gaussian random process. To this end, we take $H(z)$ to be a two-state Chebyshev filter with cutoff frequency of 0.2 times the sampling rate. The sensor noise is taken to be Gaussian and white with variance such that the SNR is 10 dB. We use the Gilbert-Elliot loss model described in Section 5. The loss probability parameters are varied with $\lambda_1 \in [0, 1]$ and $\lambda_2 = 1 - \lambda_1^2$. Under this assumption, λ_2 represents the loss probability. For each (λ_1, λ_2) pair, we run the Kalman filter on a randomly generated 1000-sample trajectory and compute the average estimation error over the trajectory. This is compared to the upper bound provided by Theorem 1. The result is plotted in Fig. 2. The estimation error is plotted relative to the norm of the signal, in dB, as a function of λ_2. As we can see, Theorem 1 in fact gives an upper bound. The LMI bound matches the simulated performance for λ_2 near 0 or near 1 and is within 2 dB otherwise.

The result of second set of computations is shown in Fig. 3. The signal model is unchanged, but now the channel is modeled as losing sensor measurements in bursts of length two. As before, for a large number of overall loss probabilities, we run the Kalman filter on a randomly generated 1000-sample trajectory and compute the average estimation error over the trajectory. This is compared to the upper bound provided by Theorem 1. The result is plotted in Fig. 3. Again, Theorem 1 gives an upper bound that matches the simulated performance very well.

As described in Section 5, various other computations could be made with no modification of the framework. This includes the possibility of variable SNR at the sensors: In certain situations, poor sensor samples are better modeled
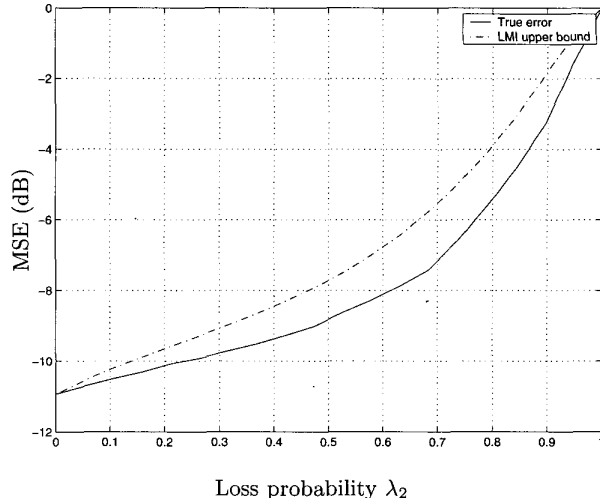

Figure 2: **Example calculations with Gilbert-Elliot loss model. The signal source is approximately low-pass with bandwidth 0.2 times the sampling frequency. As the overall sensor sample loss probability increases, the asymptotic estimation MSE increases. Theorem 1 in fact gives an upper bound that approximates the simulated performance of the optimal estimator.**

as samples with increased noise as opposed to complete erasures. This situation is easily modeled in the jump linear model (14) through the sensor noise matrix $D_{\theta(k),2}$.

8. APPLICATIONS

Multiple sensors: Accommodating multiple sensors does not require any change to the fundamental framework. Recalling the notation of Section 3, each sensor would have its own (C_s, D_s) pair and hence its own (C_2, D_2) pair. Without losses, the data from which to compute an estimate is the concatenation (stack) of $y(k)$s. With losses, different, potentially-coupled Markov chains may describe the relationships between $y_0(k)$s and $y(k)$s. The Cartesian product of the state spaces of the individual Markov chains gives an overall state space for a jump linear system. This all thus fits within our framework.

Multiple data collection nodes: Multiple data collection nodes can also be analyzed jointly without changing the basic framework. A direct application of Theorem 1 gives an upper bound to the sum of estimation error variances at the data collection nodes assuming *optimal, collaborative estimation* among the nodes. Analyzing the data collection nodes separately and jointly may thus help in understanding the merits of collaborative estimation.

Predictive quantization: When the process being monitored by the sensor is slowly-varying, the sensor data may be compressed with some form of predictive quantization before transmission to the receiver. If the prediction filter is linear and the quantization noise is modeled as additive and white, then the predictive quantizer can be incorporated into the linear system model. Specifically, the states of the linear system can be extended to include the states of the
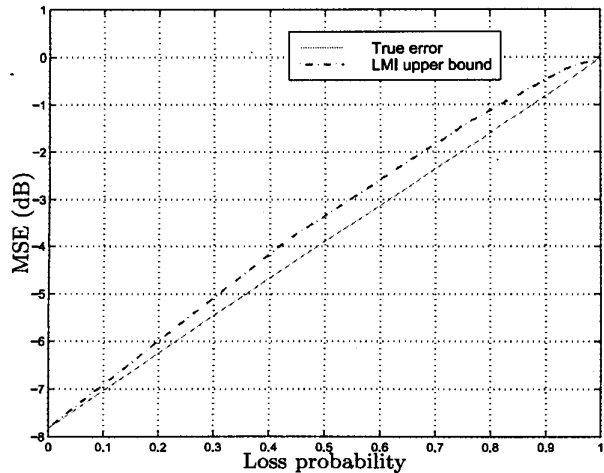
Figure 3: Example calculations with burst losses of fixed length two. Again, Theorem 1 gives an upper bound on estimation error variance that approximates the simulated performance of the optimal estimator.

prediction filter and the noise can be extended to include the quantization noise. This is explored further in [12].

Acknowledgments

9. REFERENCES

[1] C. Andrieu, M. Davy, and A. Doucet. Efficient particle filtering for jump Markov systems. application to time-varying autoregressions. *IEEE Trans. Signal Proc.*, 51(7):1762–1770, July 2003.

[2] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* SIAM, Philadelphia, PA, 1994.

[3] H. J. Chizeck, A. S. Willsky, and D. Castano. Discrete-time Markovian jump linear quadratic optimal control. *Int. J. Contr.*, 43:213–31, 1986.

[4] O. L. V. Costa. Optimal linear filtering for discrete-time Markovian jump linear systems. In *Proc. IEEE Conf. Dec. & Contr.*, volume 3, pages 2761–2762, Dec. 1991.

[5] O. L. V. Costa and M. D. Fragaso. Stability results for discrete-time linear systems with Markovian jumping parameters. *Int. J. Contr.*, 66:557–579, 1993.

[6] O. L. V. Costa and R. P. Marques. Robust H_2-control for discrete-time Markovian jump linear systems. *Int. J. Contr.*, 73(1):11–21, 2000.

[7] C. E. de Souza and M. D. Fragoso. On the existence of a maximal solution for generalized algebraic Riccati equation arising in the stochastic control. *Sys. & Contr. Letters*, 14:233–39, 1990.

[8] P. Dorato. Theoretical developments in discrete-time control. *Automatica*, 19:395–400, 1993.

[9] A. Doucet and C. Andrieu. Iterative algorithms for state estimation of jump Markov linear systems. *IEEE Trans. Signal Proc.*, 49(6):1216–1226, June 2001.

[10] A. Doucet, A. Logothetis, and V. Krishnamurthy. Stochastic sampling algorithms for state estimation of jump Markov linear systems. *IEEE Trans. Automat. Control*, 45(1):188–202, Jan. 2000.

[11] J. S. Evans and R. J. Evans. State estimation for Markov switching systems with modal observations. In *Proc. IEEE Conf. Dec. & Contr.*, pages 1688–1693, Dec. 1997.

[12] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran. Robust predictive quantization: A new analysis and optimization framework. In *Proc. IEEE Int. Symp. Inform. Th.*, Chicago, IL, June–July 2004. To appear.

[13] M. Gastpar and M. Vetterli. Scaling laws for homogeneous sensor networks. In *Proc. 41st Ann. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Oct. 2003.

[14] N. N. Krasovskii and E. A. Lidskii. Analytical design of controllers in systems with random attributes I, II, III. *Automation Remote Contr.*, 22:1021–1025, 1141–1146, 1289–1294, 1961.

[15] V. Krishnamurthy and G. G. Lin. Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime. *IEEE Trans. Inform. Th.*, 48(2):458–476, Feb. 2002.

[16] P. R. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification, and Adaptive Control.* Prentice-Hall, Englewood Cliffs, NJ, 1986.

[17] A. Logothetis and V. Krishnamurthy. Estimation maximization algorithms for MAP estimation of jump Markov linear systems. *IEEE Trans. Signal Proc.*, 47(8):2139–2156, Aug. 1999.

[18] M. A. Rami and L. El Ghaoui. LMI optimization for nonstandard Riccati equations arising in stochastic control. *IEEE Trans. Automat. Control*, 41:1666–1671, 1996.

[19] P. Seiler and R. Sengupta. Analysis of communication losses in vehicle control problems. In *Proc. IEEE Amer. Contr. Conf.*, pages 25–27, June 2001.

[20] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry. Kalman filtering with intermittent observations. In *Proc. IEEE Conf. Dec. & Contr.*, Maui, Hawaii, Dec. 2003. to appear.

[21] S. C. Smith and P. Seiler. Optimal pseudo-steady-state estimators with Markovian intermittent measurements. In *Proc. IEEE Amer. Contr. Conf.*, pages 3021–3027, May 2002.

[22] D. D. Sworder. Feedback control of a class of linear systems with jump parameters. *IEEE Trans. Automat. Control*, AC-14:60–63, 1969.

[23] W. M. Wonham. On a matrix Riccati equation of stochastic control. *SIAM J. Contr.*, 6(4):681–697, 1968.

[24] F. Zhao and L. J. Guibas, editors. *Information Processing in Sensor Networks, Second International Workshop, IPSN 2003, Palo Alto, CA, USA, April 22-23, 2003*, volume 2634 of *Lecture Notes in Computer Science.* Springer, 2003.