Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Optimal Estimation over Channels with Limits on Usage

Praveen A. Bommannavar * Tamer Başar *

* *Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 W. Main Street, Urbana, IL 61801. E-mails: bommanna@uiuc.edu, tbasar@control.csl.uiuc.edu*

Abstract: We consider several cases of a sequential estimation problem in which two decision making agents work together but with limited communication (between them) to minimize a performance criterion. One agent makes sequential observations about the state of an underlying, possibly vector valued, stochastic process for a fixed period of time. This observer agent upon observing the process decides whether or not to disclose some information about the process to the other agent, the estimator, and if *yes*, when and how. The constraint is that the observer may act only a limited, pre-specified number of times. For such problems, we develop the optimal observer-estimator policies first for the case when the source process is n^{th} order Gauss-Markov, and then for the case when the source is a vector process.

1. INTRODUCTION

The emergence of networked real-time systems during the last decade has led to a number of challenging problems in control and communications, due to nontraditional requirements on control and communication policies and architectures in such settings, such as communication and coordination of various components of what used to be a centralized operation over digital and wireless networks under bandwidth, energy, and usage constraints [1]. In this paper we address one such constraint, channel usage limitation, in the context of estimation of stochastic processes. Even though the topic of estimation of random processes using noisy state information has been thoroughly studied in the literature and has long been a fairly mature field with a complete theory [2], with the kind of nontraditional (but realistic) limitation brought about by channel usage, the emerging new classes of estimation (as well as control) problems are quite novel and challenging. Such problems were introduced recently in [3, 4, 5, 6], and we build here on this earlier work. Specifically, we are interested in the problem of estimation of a discrete-time stochastic process over a decision horizon of length N using only $M < N$ measurements. Measurement and estimation of the process are done sequentially by two different agents, called the *observer* and the *estimator*, respectively. Over the decision horizon of length N, the observer agent has exactly M opportunities to disclose some information about the process to the estimator. These *transmissions* are assumed to be error and noise free. The general problem then is to jointly determine the optimal observation and estimation policies that minimize the average estimation error between the process output (input to the observer) and its estimate (output from the estimator).

As indicated above, problems of this type have various applications. Among them are monitoring and control of wireless sensor networks, scheduling of packet transmissions over time-allocation limited channels or any other situation in which power limitation is factor [7, 8, 9].

In Sect. 2 we formally define the problem and review what is known about its solution, as pertains to this paper. Sect. 3 discusses estimating an n^{th} order Gauss-Markov scalar random sequence in which the dependence of the current state on the past reaches n time units into the past. Sect. 4 considers optimal strategies for transmitting and estimating a set of D identical, independent scalar random processes over D identical channels. Sect. 5 includes extensions in which optimal transmission/estimation of non-uniform, independent processes over non-uniform channels (using each one incurs a different cost) is developed. Simulations and examples are presented in Sect. 6, and concluding remarks are made in Sect. 7.

2. PROBLEM STATEMENT AND DEVELOPMENT

2.1 Problem Definition

Building on [3], we treat the problem of optimal estimation with limited measurements in the framework of a communication system with limited channel uses. We consider the communication system of Figure 1. The source outputs some vector data b_k for $0 \le k \le N-1$, to be communicated to a user over a channel, with the data $\{b_k\}$ generated according to some known stochastic process which may be i.i.d., or correlated as in a Markov process; further, it could be a scalar or a vector process. An encoder/observer and a decoder/estimator are placed after the source output and the channel output, respectively, to relay the data to the user optimally. Additionally, the encoder/observer may have access to a noise corrupted version of the source output: $z_k = b_k + v_k, 0 \le k \le N-1$, where v_k is a noise process, independent of b_k for all k and with no correlation across time.

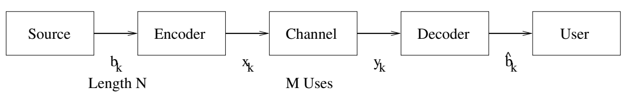

Fig. 1. Communication with limited channel use.

978-1-1234-7890-2/08/$20.00 © 2008 IFAC **10.3182/20080706-5-KR-1001.1403**

The primary constraint is that the encoder/observer can use the channel only a limited, $M < N$, number of times. For a vector process, we view the channel to be composed of several bands, each corresponding to a component of the process. All bands are used simultaneously when a transmission is made, and this counts as one channel use. We assume the channel to be memoryless and noiseless. The goal in these problems is to design an observer-estimator pair, $(\mathscr{O}, \mathscr{E})$, that will sequentially observe/encode the data measurements, z_k, and estimate/decode the channel output, y_k, while minimizing the error between the actual data, b_k, and estimated data, $\hat{b}_k$.

We define the estimation error by the standard mean square measure:

$$e_{(M,N)} = E\left\{\sum_{k=0}^{N-1}\sum_{i=1}^{D}(b_{ki} - \hat{b}_{ki})^2\right\} \tag{1}$$

where D represents the dimension of the vectors, b_k, i is the subscript denoting the component of the vector, and k is the subscript denoting the time step.

Using this framework, with the source, channel and estimation error defined, our problem is formally stated as follows: Given a source and a memoryless channel, for a given decision-horizon N, and a given number of channel uses M, what is the minimum attainable value of the estimation error $e_{(M,N)}$? This minimization is carried out over all *causal* encoder-decoder (observer-estimator) pairs.

2.2 Problem Development

In [3] a special case of the defined problem, with a zero-mean i.i.d. scalar random sequence $b_k, 0 \leq k \leq N-1$, was considered. The b_k's are assumed to have a finite second moment, σ_b^2. Let $\mathscr{B}$ denote the range of the random variable b_k. At time k, the encoder/observer makes a sequential measurement of b_k and decides whether to use one of $M < N$ channel transmissions. The channel is noiseless and even when the observer/encoder decides not to transmit, a 1-bit information may still be conveyed across the channel. Specifically, the channel input x_k belongs to the set $\mathscr{X} := \mathscr{B} \cup NT$, where NT stands for "no transmission." Let s_k denote the number of transmissions remaining at time k. Both the encoder and decoder can keep track of this by initializing $s_0 = M$ and decrementing it by 1 every time a transmission is made. The objective is to design an estimator/decoder $\hat{b}_k = \hat{\mu}_k(I_k^d), 0 \leq k \leq N-1$, based on the information I_k^d available at time k, which is a result of an outcome of decisions taken by the observer until time k. The observer's decision at time k is denoted by $x_k = \mu_k(I_k^e), 0 \leq k \leq N-1$, where I_k^e is the information available to the observer at time k. The range of $\mu_k(\cdot)$ is $\mathscr{X}$ as defined above. We assume perfect recall and have

$$I_0^e = \{(s_0, t_0); b_0\}$$
$$I_k^e = \left\{(s_k, t_k); b_0^k; x_0^{k-1}\right\}, 1 \leq k \leq N-1$$

where t_k and s_k denote respectively the number of time slots and transmissions left at time k. The channel output y_k is $y_k = x_k$ if $s_k \geq 1$, and $y_k \in \emptyset$ (no information) if $s_k = 0$. So we can write the information I_k^d available to the estimator at time k as $I_k^d = \{(s_k, t_k); y_0^k\}, 0 \leq k \leq N-1$.

In view of (1), we want to find an admissible policy that minimizes the N stage estimation error. In [3], the problem was also considered for random processes which are Gauss-Markov in nature.

To summarize, the problem is sequential in nature: at time k, b_k is observed, then the observer decides whether or not to transmit some data. The estimator makes an estimate, incurring some (possibly zero) cost and we move on to the next time $k+1$, and so on.

2.3 Relevant Results

Only the case $M < N$ is considered, because otherwise the result is trivial. For the zero mean i.i.d. problem it is shown in [3] that the optimal estimator is

$$\hat{\mu}_k(I_k^d) = \begin{cases} E\{b_k\} = 0 & \text{if } s_k = 0 \\ E\{b_k | (s_k, t_k); x_k\} & \text{if } s_k > 0 \end{cases}$$

It is also shown that the observer's decision to use the channel to transmit a source measurement or not is based purely on the current observation, b_k, and its past actions only through (s_k, t_k). As a result, the optimal observer policy must be of the form:

$$\mu_k(I_k^e) = \begin{cases} b_k & \text{if } b_k \in \mathscr{F}_{(s_k,t_k)} \\ NT & \text{if } b_k \in \mathscr{F}^c_{(s_k,t_k)} \end{cases}$$

where $\mathscr{F}_{(s_k,t_k)}$ is a measurable set on $\mathscr{B}$, and is a function of (s_k, t_k). The complement of this set is with respect to $\mathscr{B}$: $\mathscr{F}^c_{(s_k,t_k)} = \mathscr{B} \backslash \mathscr{F}_{(s_k,t_k)}$.

Two salient points emerge in these results. The first is that the optimal policy for the estimator/observer pair is to decouple the past from the present (and future). A related point is that the error incurred after time k, i.e.,

$$e_k = E\left\{\sum_{n=k}^{N-1}\sum_{i=1}^{D}(b_{ni} - E\{b_n | (s_n, t_n); x_n\}_i)^2\right\}$$

also depends only on b_k and (s_k, t_k).

Let $(s_k, t_k) = (s, t)$, and $e^*_{(s,t)}$ denote the optimal value of the estimation error when the decision horizon is of length t and the observer is limited to s channel uses, where $s \leq t$. Supposing the value of b_k is unknown, the estimation error can be expressed recursively using a dynamic programming (DP) approach [10]: depending on what b_k turns out to be, the remaining $(t-1)$-stage estimation error is either $e^*_{(s-1,t-1)}$ or $e^*_{(s,t-1)}$. In the $D = 1$ case, therefore, we may write

$$e^*_{(s,t)} = \min_{\mathscr{F}_{(s,t)}} \left\{ e^*_{(s-1,t-1)} \int_{b \in \mathscr{F}} f(b)db + e^*_{(s,t-1)} \cdot \right.$$
$$\left. \int_{b \in \mathscr{F}^c} f(b)db + \int_{b \in \mathscr{F}^c} \left[b - E\{b | b \in \mathcal{F}^c\}\right]^2 f(b)db \right\}$$

The second important point is that the optimization over the observer policies is equivalent to optimization over the sets $\mathscr{F}_{(s_k,t_k)}$ for all k such that $\max\{0, M\} \leq s_k \leq \min\{t_k, M\}$. The approach taken in [3] as well as in this paper is to restrict our search to sets $\mathscr{F}$ that are in the form of simple intervals, i.e., $\mathscr{F}^c_{(s,t)} = [\alpha_{(s,t)}, \beta_{(s,t)}]$. By rewriting the above integral in terms of only $\mathscr{F}^c_{(s,t)}$, we can differentiate with respect to the endpoints of the integral to obtain their optimal values (in terms of (s,t)).

Similar results hold for a slight variant of this problem in which the source process is Markov, as indicated above. Intuitively we can understand that the past and present will not be decoupled as opposed to the i.i.d. case but will be linked by a variable connecting the two. To make this formal, in [3] it was established that for the Gauss-Markov problem indicated above that the optimal estimator is

$$\hat{\mu}_k(I^d_{N-t}) = \begin{cases} E\{b_{N-t}\} = A^r b_{N-t-r} & \text{if } s = 0 \\ E\{b_{N-t}|(r,s,t); x_{N-t}, x_{N-t-r}\} & \text{if } s > 0 \end{cases}$$

It was also established that the observer's decision whether to use the channel to transmit a source measurement or not is based on the current observation, b_k, and its past actions only through (r, s, t). As a result, the optimal observer policy must be of the form:

$$\mu_k(I^e_{N-t}) = \begin{cases} b & \text{if } b \in \mathscr{F}_{(r,s,t)} \\ NT & \text{if } b \in \mathscr{F}^c_{(r,s,t)} \end{cases}$$

Three main problems are solved for the 1-dimensional case in [3]. First an error recursion is developed and decision regions are derived for the case in which b_k is a Gaussian, i.i.d. process. Then similar results are derived for the case when measurements are corrupted by noise. Finally, the recursion and decision region is found for the case when b_k is a Gauss-Markov process.

3. HIGHER ORDER GAUSS-MARKOV CASE

In this section, we discuss the case when the source process is a multi-step Markov process, i.e. the dependence of the current state is on more than one previous state; in other words, there is further depth into the past:

$$b_{k+1} + a_0 b_k + ... + a_{n-1} b_{k-n+1} = w_k$$

where w_k is zero mean, i.i.d. Gaussian with variance σ_w^2. Additionally, each transmission is able to make use of n channels, each of which is noiseless. Just as in the Gauss-Markov case, the past and the present are linked. In order to utilize the method of Sect. IV of [3], we want to keep track of one variable that represents this link so that as before, the observer estimator pair has to keep track of (r_k, s_k, t_k). Here r is the number of time units since the last transmission, s is the number of transmissions left, and t is the number of time units left. To this end, we introduce the following framework:

$$\eta_k = (\, b_{k-n+1} \ \dots \ b_k \,)^T$$

$$\eta_{k+1} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-1} \end{pmatrix} \eta_k + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} w_k$$

$$y_k = (\, 0 \ \dots \ 0 \ 1 \,)\, \eta_k =: C\eta_k$$

where we shall henceforth denote the system matrix premultiplying η_k by A. In this new framework, we consider the problem of optimally transmitting the vector η_k, and also note that $y_k = b_k$. Since optimal estimation of the vector is equivalent to optimal estimation of each of the components, we can determine the best estimate for η_k and then use this to get the best estimate for b_k. Also, the results of [3] tell us that if a transmission is to be made at time k, the transmission should be η_k itself. Note also that to send the most information about current and future values of the process, we must transmit the most recent measurements that determine the future. In our case, that corresponds to vector η_k. Hence, use of the state space model is justified. Reasoning in the same way as in Sect. IV of [3], for $s \geq 1$ the optimal estimator has the form

$$\hat{\mu}((r,s,t); \eta_{N-t}) = \begin{cases} \eta_{N-t} & \text{if } y_{N-t} \in \mathscr{F}_{(r,s,t)} \\ E\{\eta_{N-t}|y_{N-t} \in \mathscr{F}^c_{(r,s,t)}\} & \text{if } y_{N-t} \in \mathscr{F}^c_{(r,s,t)} \end{cases}$$

We take the last component of the vector $\hat{\mu}$ to get the desired estimate, y_{N-t}. The quantity $E\{\eta_{N-t}|y_{N-t} \in \mathscr{F}^c_{(r,s,t)}\}$ can be computed using the Kalman Filter. We will denote this quantity, the best estimate for η_{N-t}, by $\hat{\eta}_{N-t|N-t-r}$. First we note that $\hat{\eta}_{N-t|N-t-r} = A^r \eta_{N-t-r}$, where η_{N-t-r} is known because it was transmitted r time units ago. To determine the optimal observer structure and error recursion, we now make use of the following information, where $\mathcal{N}$ denotes the Gaussian distribution:

$$b_{N-t} = y_{N-t} \sim \mathcal{N}(CA^r \eta_{N-t-r}, C\Sigma_{N-t|N-t-1}C^T)$$

Here $\Sigma_{N-t|N-t-1}$ is the error matrix associated with estimating the vector η_{N-t} using information until time $N-t-1$. As verified below, this relevant information consists of the last transmission and the number of time units since that last transmission was made.

Claim 1. $\Sigma_{N-t|N-t-1}$ can be expressed in terms of A, r and σ_w^2.

This can be seen by analyzing the Kalman Filter expressions. Carrying out the necessary calculations reveals that

$$\Sigma_{N-t|N-t-1} = \sum_{k=1}^{r} A^{k-1} \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_w^2 \end{pmatrix} (A^T)^{k-1}$$

Next we derive the error recursion:

$$e^*_{(r,s,t)} = \min_{\mathscr{F}_{(r,s,t)}} \Bigg(e^*_{(1,s-1,t-1)} P_1 + e^*_{(r+1,s,t-1)} P_2 + \int_{b_{N-t} \in \mathscr{F}^c_{(r,s,t)}} (b_{N-t} - \hat{\mu}_{N-t}((r,s,t); \hat{\eta}_{N-t}))^2 \cdot f_{b_{N-t}}(b_{N-t}) db_{N-t} \Bigg)$$

where $P_1 = Pr[b_{N-t} \in \mathscr{F}_{(r,s,t)}]$ and $P_2 = Pr[b_{N-t} \in \mathscr{F}^c_{(r,s,t)}]$. Clearly $P_1 + P_2 = 1$. Substituting known quantities, we can simplify the above expression:

$$e^*_{(r,s,t)} = \min_{\mathscr{F}^c_{(r,s,t)}} \Bigg(e^*_{(1,s-1,t-1)} - (e^*_{(1,s-1,t-1)} + e^*_{(r+1,s,t-1)}) \cdot \int_{b_{N-t} \in \mathscr{F}^c_{(r,s,t)}} f_{b_{N-t}}(b_{N-t}) db_{N-t} + \int_{b_{N-t} \in \mathscr{F}^c_{(r,s,t)}} (b_{N-t} - CA^r \eta_{N-t-r})^2 f_{b_{N-t}}(b_{N-t}) db_{N-t} \Bigg)$$

We conduct this optimization by restricting $\mathscr{F}^c_{(r,s,t)}$ to be in the form of a simple interval: $\mathscr{F}^c_{(r,s,t)} = [\alpha_{(r,s,t)}, \beta_{(r,s,t)}]$. We assume the simplifying symmetry $\alpha_{(r,s,t)} + \beta_{(r,s,t)} = 2CA^r \eta_{N-t-r}$. The result of simply differentiating and obtaining optimal thresholds is

$$\alpha = CA^r \eta_{N-t-r} - \sqrt{e^*_{(1,s-1,t-1)} - e^*_{(r+1,s,t-1)}}$$
$$\beta = CA^r \eta_{N-t-r} + \sqrt{e^*_{(1,s-1,t-1)} - e^*_{(r+1,s,t-1)}}$$

Substitution of these optimum thresholds yields:

$$\begin{aligned} e^*_{(r,s,t)} &= e^*_{(1,s-1,t-1)} - \\ &\left[e^*_{(1,s-1,t-1)} - e^*_{(r+1,s,t-1)} - C\Sigma_{N-t|N-t-1}C^T\right] \cdot \\ &\left[2\Phi\left(\sqrt{\frac{e^*_{(1,s-1,t-1)} - e^*_{(r+1,s,t-1)}}{C\Sigma_{N-t|N-t-1}C^T}}\right) - 1\right] - \\ &\frac{2}{\sqrt{2\pi}}\sqrt{C\Sigma_{N-t|N-t-1}C^T} \cdot \\ &\sqrt{e^*_{(1,s-1,t-1)} - e^*_{(r+1,s,t-1)}}\, e^{-\frac{e^*_{(1,s-1,t-1)} - e^*_{(r+1,s,t-1)}}{2C\Sigma_{N-t|N-t-1}C^T}} \end{aligned}$$

This recursion is defined for $r \geq 1$ and $0 \leq s \leq t$ with boundary conditions given by

$$e^*_{(r,t,t)} = 0$$
$$e^*_{(r,0,t)} = \sum_{l=r}^{r+t-1} \sum_{k=1}^{l} CA^{k-1} \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_w^2 \end{pmatrix} (A^T)^{k-1} C^T$$

For $n = 1$, these results reduce to those of Sect. IV of [3], which corresponds to a one step Markov process for which we have one channel available for transmission. The generalization allows for an n step Markov process with the caveat that each transmission allows us to send a vector of n dimensions. If we have less than n channels, Claim 1 is no longer valid, and we cannot proceed with a DP approach. Note also that all optimal transmission/estimation problems dealing with a Markov process must begin with a transmission at the first time step.

4. DUPLICATION ACROSS D CHANNELS

Here we discuss a series of related problems. Consider a scalar random process, for example a Gaussian i.i.d. process with bounded variance. Now suppose the observer makes measurements of D such i.i.d. processes. Optimal policies for the observer/estimator pair, $(\mathscr{O}, \mathscr{E})$, can be deduced by modeling the D processes as independent components of a vector random process. We first carry out the formal derivations for the Gaussian i.i.d. case. Then we reconsider the problem if each of the D processes is again Gaussian i.i.d., but also corrupted by noise. Finally we suppose that each of the processes is Gauss-Markov or n^{th} order Gauss-Markov. Note that the solution is not trivial - to minimize the error criterion, we should consider all components of the vector, i.e. applying the known solution in one dimension to the first component will not be optimal in general. Also, simply considering sequences that are D times as long as the original sequence (with D times the number of transmissions) will not work. There is the additional constraint that if a vector is sent at a particular time, we must send all components at once.

This problem can be applicable to network problems in which one may have a centralized scheduler at the source.

4.1 i.i.d. Gaussian case

Suppose each of our D identical and independent zero mean Gaussian processes has variance σ_b^2. Then we may represent the processes, $b^1, ..., b^D$, as components of a vector process b with statistics $b \sim N(0, \Sigma_b)$ where $\Sigma_b = \sigma_b^2 I$, since the processes are independent and of equal variance. We denote the corresponding pdf by $f(b^1, ..., b^D)$. From the previous results, we see that the optimal estimator is

$$\hat{\mu}_k((s,t); x_k) = \begin{cases} x_k & \text{if } b_k \in \mathscr{F}_{(s,t)} \\ E\{b_k | b_k \in \mathscr{F}^c_{(s,t)}\} & \text{if } b_k \in \mathscr{F}^c_{(s,t)} \end{cases}$$

where x_k is now a vector version of the same variable in Fig. 1 (it represents what is received by the estimator). As before, we assume a certain degree of symmetry in the decision region $\mathscr{F}^c_{(s,t)}$. We allow the D dimensional region to be in the shape of a symmetrical hypercube. That is, for each axis i, we allow $\mathscr{F}^c_{(s,t),i} = [-\alpha_i, \alpha_i]$. Proceeding as above by developing the error recursion, we obtain:

$$\begin{aligned} e^*_{(s,t)} = e^*_{(s-1,t-1)} + \min_{\mathscr{F}^c_{(s,t)}} \Bigg(&- (e^*_{(s-1,t-1)} - e^*_{(s,t-1)}) \cdot \\ &\int_{-\alpha_1}^{\alpha_1} \cdots \int_{-\alpha_D}^{\alpha_D} f(b^1, \dots, b^D) db^1 \dots db^D \\ &+ w_1 \int_{-\alpha_1}^{\alpha_1} \cdots \int_{-\alpha_D}^{\alpha_D} (b^1)^2 f(b^1, \dots, b^D) db^1 \dots db^D + \dots \\ &+ w_D \int_{-\alpha_1}^{\alpha_1} \cdots \int_{-\alpha_D}^{\alpha_D} (b^D)^2 f(b^1, \dots, b^D) db^1 \dots db^D \Bigg) \end{aligned} \quad (2)$$

where for now, $w_1, ..., w_D$ are equal to 1. We want to minimize the quantity globally with respect to $\mathscr{F}^c_{(s,t)}$. We first take partial derivatives with respect to α_k:

$$\begin{aligned} \frac{\partial e^*_{(s,t)}}{\partial \alpha_k} = \frac{2e^{\frac{-\alpha_k^2}{2\sigma_b^2}}}{(2\pi\sigma_b^2)^{D/2}} \prod_{j=1, j\neq k}^{D} \left(\int_{-\alpha_j}^{\alpha_j} e^{\frac{-(b^j)^2}{2\sigma_b^2}} db^j \right) \cdot \\ \Bigg\{ - \left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)} \right) + \alpha_k^2 + \\ \sum_{i=1, i\neq k}^{D} \frac{\int_{-\alpha_i}^{\alpha_i} (b^i)^2 e^{\frac{-(b^i)^2}{2\sigma_b^2}} db^i}{\int_{-\alpha_i}^{\alpha_i} e^{\frac{-(b^i)^2}{2\sigma_b^2}} db^i} \Bigg\} \end{aligned}$$

Upon inspecting this quantity, we arrive at the following main result of this subsection:

Theorem 2. $e^*_{(s,t)}$ has a unique global minimizing decision region, $\mathscr{F}^c_{(s,t)}$ when restricted to sets of the form $\mathscr{F}^c_{(s,t),i} = [-\alpha_i, \alpha_i]$ with $1 \leq i \leq D$ and $\alpha_i > 0$. This minimizer is given by $\alpha_i = \alpha$, where $\alpha > 0$ satisfies

$$\begin{aligned} 0 = &- \left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)} \right) + \alpha^2 \\ &+ (D-1) \left(\sigma_b^2 - \frac{2\sigma_b \alpha e^{\frac{-\alpha^2}{2\sigma_b^2}}}{\sqrt{2\pi}(2\Phi\left(\frac{\alpha}{\sigma_b}\right) - 1)} \right), \end{aligned}$$

where Φ is the standard Gaussian error function.

Proof: Using the usual rules of integration for Gaussian pdf's, it is easily shown that the given decision region is indeed a stationary point. This is done by noting that the choice causes all partial derivatives to be equal to zero. In the expression above, we focus on the part inside the curly braces (the part multiplying it is always positive). Substituting $\alpha_i = \alpha$ in this expression, we get exactly the expression on the RHS above. To show that this admits a unique solution, and that this solution is the unique minimum sought, we proceed as follows. First note at the point $\alpha_k = 0$, the expression approaches a strictly negative value. However, as α_k becomes large, the expression becomes strictly positive. Finally, at any point where the expression is zero, moving an arbitrary amount to the right increases the quantity and moving to the left decreases the quantity. Hence we see that a solution exists and is unique, with all α_k's the same and positive. $\diamond$

We may now wrap up the problem by rewriting the error recursion in terms of α. This is done, as above, by substitution of the optimal thresholds into the recursion:

$$e^*_{(s,t)} = e^*_{(s-1,t-1)} - (e^*_{(s-1,t-1)} - e^*_{(s,t-1)})\cdot \left(2\Phi\left(\frac{\alpha}{\sigma_b}\right)-1\right)^D + D\left(2\Phi\left(\frac{\alpha}{\sigma_b}\right)-1\right)^{D-1}\cdot \left[\sigma_b^2\left(2\Phi\left(\frac{\alpha}{\sigma_b}\right)-1\right) - \frac{2\sigma_b\alpha}{\sqrt{2\pi}}e^{-\frac{\alpha^2}{2\sigma_b^2}}\right]$$

The recursion is defined for $0 \le s \le t$ with boundary conditions given by $e^*_{(t,t)} = 0$ and $e^*_{(0,t)} = D\sigma_b^2 t$.

4.2 Noisy i.i.d. Gaussian case

Suppose each of the D i.i.d. zero mean Gaussian processes has variance σ_b^2 and the observations are corrupted by i.i.d. zero mean Gaussian noise of variance σ_v^2. We may write: $z_k = b_k + v_k$ where b_k is defined as in the previous subsection, v_k is a mean zero Gaussian noise vector with covariance matrix $\sigma_v^2 I$ and b_k and v_k are independent. The following distributions are also known:

$$f_{z_k} \sim N\left(0, (\sigma_b^2+\sigma_v^2)I\right)$$
$$f_{b_k|z_k} \sim N\left(\frac{\sigma_b^2}{\sigma_b^2+\sigma_v^2}z_k, \left(\frac{\sigma_b^2\sigma_v^2}{\sigma_b^2+\sigma_v^2}\right)I\right)$$

After reasoning in the usual manner, we obtain the optimal estimator and observer structures:

$$\hat{\mu}_k((s,t);x_k) = \begin{cases} \frac{\sigma_b^2}{\sigma_b^2+\sigma_v^2}x_k & \text{if } b_k \in \mathscr{F}_{(s,t)} \\ E\{b_k | z_k \in \mathscr{F}^c_{(s,t)}\} & \text{if } b_k \in \mathscr{F}^c_{(s,t)} \end{cases}$$

$$\mu_k((s,t);z_k) = \begin{cases} \hat{z}_k & \text{if } z_k \in \mathscr{F}_{(s,t)} \\ NT & \text{if } z_k \in \mathscr{F}^c_{(s,t)} \end{cases}$$

The optimal decision regions are given by $\mathscr{F}^c_{(s,t),i} = [-\frac{\sigma_b^2+\sigma_v^2}{\sigma_b^2}\alpha, \frac{\sigma_b^2+\sigma_v^2}{\sigma_b^2}\alpha]$ where the subscript i is used to index the components of our vector and $\alpha > 0$ is the solution to

$$0 = -\left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)}\right) + \alpha^2 + (D-1)\left(\frac{\sigma_b^2}{\sigma_b^2+\sigma_v^2}\right)\cdot \left(\sigma_b^2 - \frac{2\sqrt{\sigma_b^2+\sigma_v^2}\,\alpha e^{\frac{-\sqrt{\sigma_b^2+\sigma_v^2}}{2(\sigma_b^2)^2}\alpha^2}}{\sqrt{2\pi}(2\Phi\left(\frac{\sqrt{\sigma_b^2+\sigma_v^2}}{\sigma_b^2}\alpha\right)-1)}\right)$$

Finally, the error recursion is given by

$$e^*_{(s,t)} = e^*_{(s-1,t-1)} + D\frac{\sigma_b^2\sigma_v^2}{\sigma_b^2+\sigma_v^2} - (e^*_{(s-1,t-1)} - e^*_{(s,t-1)})\cdot \left(2\Phi\left(\frac{\alpha}{\tilde{\sigma}}\right)-1\right)^D + D\left(2\Phi\left(\frac{\alpha}{\tilde{\sigma}}\right)-1\right)^{D-1}\cdot \left[\tilde{\sigma}^2\left(2\Phi\left(\frac{\alpha}{\tilde{\sigma}}\right)-1\right) - \frac{2\alpha}{\sqrt{2\pi}\tilde{\sigma}}e^{-\frac{-\alpha^2}{2\tilde{\sigma}^2}}\right]$$

where $\tilde{\sigma} = \frac{\sigma_b^2}{\sqrt{\sigma_b^2+\sigma_v^2}}$. The recursion is defined for $0 \le s \le t$ with boundary conditions given by $e^*_{(t,t)} = D\frac{\sigma_b^2\sigma_v^2}{\sigma_b^2+\sigma_v^2}t$ and $e^*_{(0,t)} = D\sigma_b^2 t$. This reduces to the result of the previous subsection when the noise variance is taken to zero.

5. NON-UNIFORM VARIANCES AND NON-UNIFORM CHANNEL COSTS

We now consider a generalization of the given problem. Suppose that over each of D channels we have a mean zero Gaussian process b_i but each process has a possibly different variance, σ_{bi}^2 for $1 < i < D$. Suppose also that we wish to weight the cost of using each channel differently, that is, the cost of using channel i is w_i. In the above problems, we took $w_i = 1$ for all i. The problem is developed in the same way as in subsection 4.1. We again assume a certain degree of symmetry in the decision region $\mathscr{F}^c_{(s,t)}$ and allow the D dimensional region to be in the shape of a symmetrical hypercube. That is, for each axis i, we allow $\mathscr{F}^c_{(s,t),i} = [-\alpha_i, \alpha_i]$. Proceeding as above by developing the error recursion, we obtain (2) (this time weighting the channel costs, and again using f for the corresponding joint pdf). Taking partial derivatives gives

$$\frac{\partial e^*_{(s,t)}}{\partial \alpha_k} = \frac{2e^{\frac{-\alpha_k^2}{2\sigma_{bk}^2}}}{(2\pi)^{D/2}\sigma_{b1}\ldots\sigma_{bD}}\prod_{j=1,j\neq k}^{D}\left(\int_{-\alpha_j}^{\alpha_j} e^{\frac{-(b^j)^2}{2\sigma_{bj}^2}}db^j\right)\cdot \left\{-\left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)}\right) + w_k\alpha_k^2 + \sum_{i=1,i\neq k}^{D}\frac{w_i\int_{-\alpha_i}^{\alpha_i}(b^i)^2 e^{\frac{-(b^i)^2}{2\sigma_b^2}}db^i}{\int_{-\alpha_i}^{\alpha_i}e^{\frac{-(b^i)^2}{2\sigma_b^2}}db^i}\right\}$$

This implies that unlike before, we may not take $\alpha_i = \alpha$ for all i. Rather, we are forced to numerically solve a system of equations. These equations are obtained by setting the expression in the braces above to zero for each k, $1 \le k \le D$. In the case of D = 2, they may be decoupled easily. In other cases, the numerical computation gets increasingly complex. In the fashion of Theorem 2, it can be shown that a unique solution exists to the desired minimization (and, thus, to these equations). After solving for the thresholds α_i, we can plug them back into the recursion to obtain:

$$e^*_{(s,t)} = e^*_{(s-1,t-1)} - (e^*_{(s-1,t-1)} - e^*_{(s,t-1)})\cdot \prod_{i=1}^{D}\left(2\Phi\left(\frac{\alpha_i}{\sigma_{bi}}\right)-1\right)\left[\prod_{i=1}^{D}\left(2\Phi\left(\frac{\alpha_i}{\sigma_{bi}}\right)-1\right)\right] + \sum_{j=1}^{D} w_j\left(\sigma_{bj}^2 - \frac{2\sigma_{bj}\alpha_j e^{-\frac{\alpha_j^2}{2\sigma_{bj}^2}}}{\sqrt{2\pi}\left(2\Phi\left(\frac{\alpha_j}{\sigma_{bj}}\right)-1\right)}\right)$$

The recursion is defined for $0 \leq s \leq t$ with boundary conditions given by $e^*_{(t,t)} = 0$ and $e^*_{(0,t)} = t\sum_{i=1}^{D} w_i \sigma_{bi}^2$. Note that this development simplifies to the duplication across D channels case of subsection 4.1 when the weights w_i are taken to be one and when the variances σ_{bi} (and subsequently decision thresholds α_i) are independent of i.

6. SIMULATIONS AND EXAMPLES

6.1 A two step noiseless scalar Gauss-Markov process

Consider the following random process

$$b_{k+1} - 0.5b_k - 0.25b_{k-1} = w_k$$

where w_k is i.i.d. Gaussian with $\sigma_w^2 = 1$. This is a stable process, with the eigenvalues of the 2-dimensional matrix of its state space representation being 0.8431 and -0.5931. Implementation of the algorithm for an event horizon of $N = 30$ and varying M over $[0, 30]$ yields an error curve resembling Fig. 5 from [3]. See the left side of Fig. 2 below.

6.2 Transmitting three i.i.d. Gaussian processes

Suppose we have three random processes, all mean zero i.i.d. Gaussian with variance $\sigma_b^2 = 0.8$. Say that we are interested in the process over a horizon of $N = 40$ time steps and the observer, as defined above, is allowed $M < 40$ channel uses to transmit this information to an estimator agent. A channel use means transmitting the state of all three processes simultaneously. The right side of Fig. 2 shows the corresponding optimal error vs. number of allowed channel uses graph.

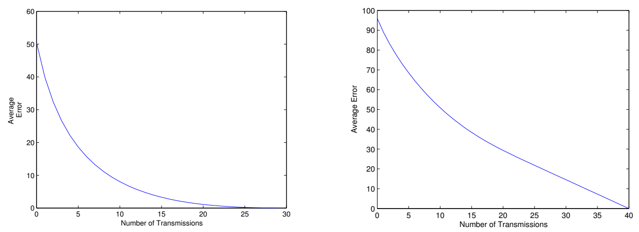

Fig. 2. Optimal 40-stage estimation error vs. the number of allowed channel uses under two scenarios.

Now suppose that rather than having three identical processes, we have 30 such processes. Plotting the graph again yields a much straighter curve, which seems to indicate a trend. As more processes are considered, the curve becomes a straight line connecting the points (0,32P) and (40,0) where P is the number of processes. This is a topic that warrants further investigation.

7. CONCLUSION

We have obtained optimal transmission/estimation strategies for a general n-step Gauss-Markov process with limited channel usage. We have done this by considering a state space model and transmitting a vector process over the channel. We have also obtained the policies for sending vectors over a channel with limited transmissions, as long as the components of the vector are independent.

We can generalize the results of 4.1 to the n^{th}-order Gauss-Markov case where such a process is duplicated over D channels (identical and independent as in the 0^{th} order case). The results have not been included due to space limitations. The form of the solution will be similar to the developments of 4.1. An important difference surfaces, however: the decision region is not centered around the origin, but around the most likely estimate of the process.

A number of open questions remain. An important problem is to determine optimal strategies when an n^{th}-order process is to be transmitted over fewer than n channels. Another interesting case is to consider systems in which there is a Markov dependency as well as observation noise. Additionally, both [3] and this paper restrict the analysis to symmetric decision regions; a formal proof needs to be obtained to show that there is no loss of optimality in doing this. One might also consider optimal control problems in which actuators have limited transmissions to the plant, and when these transmissions have limited reliability, along the direction of the work in [5] and [11].

ACKNOWLEDGEMENTS

Research reported here was supported in part by a National Physical Science Consortium (NPSC) fellowship, and in part by NSF Grant ANI-031976.

REFERENCES

[1] P.J. Antsaklis and J. Baillieul, (Editors). Technology of Networked Control Systems, Special Issue of Proceedings of the IEEE, 95(1), January 2007.

[2] B.D. Anderson and J.B. Moore. *Optimal Filtering.* Prentice-Hall, Inc., Englewood Cliffs, NJ. 1979.

[3] O.C. Imer and T. Başar. "Optimal estimation with limited measurements," Proc. of 44th IEEE CDC and ECC, Seville, Spain. 2005.

[4] O.C. Imer. *Optimal Estimation and Control under Communication Network Constraints.* Ph.D. Dissertation, University of Illinois at U-C. 2005.

[5] O.C. Imer and T. Başar. "Optimal control with limited controls," In: Proc. of American Control Conference (ACC), Minneapolis, MN, 2006.

[6] O.C. Imer and T. Başar. "Wireless sensing with power constraints," In: Advances in Control Theory and Applications, C. Bonivento, A. Isidori, L. Marconi, and C. Rossi (Eds.), Lecture Notes in Control and Information Sciences, vol. 353, pp. 129-160, Springer Verlag, 2007.

[7] J.A. Gutierrez, E.H. Callaway and R. L. Barrett Low Rate Wireless Personal Area Networks: Enabling Wireless Sensors with IEEE 802.15.4 IEEE Press, New York, NY. 2003.

[8] N.V. Stralen, O.C. Imer, R. Mitchell, S. Evans, and S. Iyer. A multiband random access messaging protocol. Proc. MILCOM, Washington, D.C. 2006.

[9] J. Baillieul and P.J. Antsaklis. "Control and communication challenges in networked real-time systems," Proceedings of the IEEE, 95(1): 9 - 27, January 2007

[10] D.P. Bertsekas, *Dynamic Programming and Optimal Control.* Belmont, MA: Athena Scientific, 1995

[11] O.C. Imer, S. Yüksel, and T. Başar, "Optimal control of LTI systems over communication networks," Automatica, 42(9): 1429-1440, September 2006.