2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Optimal Estimation of Multidimensional Data with Limited Measurements

W. MacKunis*, J. W. Curtis†, and P. E. K. Berg-Yuen†

*Department of Physical Sciences, Embry Riddle Aeronautical University, Daytona Beach, FL 32114

†Munitions Directorate, Air Force Research Laboratory, Eglin AFB, 32542

email: william.mackunis@erau.edu, jess.curtis@eglin.af.mil, piagreg@cox.net

Abstract—**Recent results indicate how to optimally schedule transmissions of a measurement to a remote estimator when there are limited uses of the communication channel available. The resulting optimal encoder and estimation policies solve an important problem in networked control systems when bandwidth is limited. Previous results were obtained only for scalar processes, and the previous work was unable to address questions regarding informational relevance. We extend the state-of-the art by treating the case where the source process and measurements are multi-dimensional. To this end, we develop a non-trivial re-working of the underlying proofs. Specifically, we develop optimal encoder policies for Gaussian and Gauss-Markov measurement processes by utilizing a measure of the informational value of the source data. Explicit expressions for optimal hyper-ellipsoidal regions are derived and utilized in these encoder policies. Interestingly, it is shown in this paper that analytical expressions for the hyper-ellipsoids exist only when the state's dimension is even; in odd dimensions (as in the scalar case) the solution requires a numerical look up (e.g., use of the erf function). We also have extended the previous analyses by introducing a weighting matrix in the quadratic cumulative cost function, whose purpose is to allow the system designer to designate which states are more important or relevant to total system performance.**

I. INTRODUCTION[1]

The current technological advances in circuit miniaturization are driving down the cost of producing many sensors, especially cameras and other EO imaging sensors; this makes it feasible to deploy numerous high-resolution sensors to provide feedback for some control system. One of the many emerging challenges faced by modern feedback system designers is that such sensors are often not collocated with the plant to be controlled. This leads to difficult questions regarding bandwidth utilization, time-delays, and other effects that have been somewhat addressed in the networked control systems literature (e.g., see [1], [2], [3]).

A central issue in such information-rich problems is that of deciding which data is valuable enough to transmit (and thus consume relatively scarce bandwidth resources) and which data can be safely discarded. The seminal work in [4] addressed a version of this problem, where they assumed that only one sensor among a set of multiple sensors could be used at any given time. They proved a separation property between the optimal plant control policy and the measurement control policy. The measurement control problem, which is the sensor scheduling problem, was cast as a nonlinear deterministic control problem and shown to be solvable by a tree-search in general. It was proven that if the decision to choose a particular sensor rests with the estimator, an open-loop selection strategy is optimal for a cost based on the estimate error covariance [5]. Forward dynamic programming (DP) and a gradient method were proposed for this purpose.

In this paper, we address the issue of informational value in a networked estimation and control context. In these fields there has been a traditional bias towards using the entropic formulation [6], [7] as a standard measure of information, though there has recently been increasing interest in alternative informational value metrics for cooperative and networked estimation [8] and control. In [9] the authors examine the optimal communication policy of an observer who is observing a random process and who must decide whether to send observations across a communication channel to an estimator. They discover jointly optimal policies for the observer and the estimator so as to minimize the mean-square estimation error of the observer in the case where the observer is limited in the number of times that it can transmit. A method is presented in [9] to compute the optimal transmission policy off-line via DP. A very similar problem was also treated in [10], where the optimal policy involves transmitting a measurement only if it lies outside some symmetric region centered around the mean value of the observed process. Both papers treat only the scalar case, and both papers propose solutions of an optimization problem where the objective function considers only estimation and communication errors.

The constrained uses of the communication channel necessitate information arbitrage, and the sensing agent must decide which measurement will be of most value to the estimator. The contribution of this paper is development which illustrates how observation and estimation techniques can be designed for optimal estimation of multi-dimensional data over a communication channel with limited uses. This entails a non-trivial re-working of the proofs found in [9]. Moving beyond a scalar problem is essential for exploring informational relevance issues, because it forces the observer to consider which elements of the observation vector may be most relevant to some decision-maker (controller) who is being fed information by the estimator.

An important aspect of our extension to multiple dimensions

[1]This research is supported in part by a grant from the National Research Council.

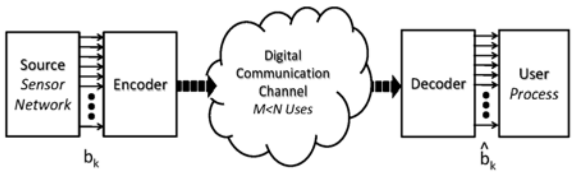

Fig. 1. Communication system with limited transmission channel uses.

is that, in the present work, the objective function can entertain notions of informational relevance through a cost-weighting matrix. In multiple dimensions one is forced to grapple with the fact that some states might be much more valuable to know precisely than others, and this leads to interesting future questions regarding how control systems being fed by this estimator might influence the choice of weights in the observer/estimator policy.

A. Problem Statement

The problem of optimal estimation of data based on limited measurements will be addressed for the case where the source data is multidimensional. The problem will be framed in the context of communication of data across a channel with limited uses as depicted in Fig. 1. By extending a technique similar to that presented in [2], value of information (VOI)-based encoder and decoder policies will be utilized for optimal sequential estimation of n dimensional data communicated over a channel with limited uses. Specifically, a source sequentially generates data $b_k \in \mathbb{R}^n$ over an N-step decision horizon $0 \leq k \leq N-1$, which must be transmitted over a channel. The data b_k are generated according to some a priori known stochastic process (e.g., an independent identically distributed (I.I.D.) Gaussian random process or a correlated Gauss-Markov process). An encoder/observer is placed at the source output, and a decoder/estimator is placed at the channel output. Observer and estimator policies are utilized to optimize the accuracy of the communication system in the presence of limited channel uses.

The communication channel is restricted such that it can only be accessed for $M < N$ transmissions. The objective is to design observer and estimator policies that minimize the error between the source data b_k and its estimate $\hat{b}_k$ over the N-step decision horizon. At each time step k, the number of remaining time steps is denoted $1 \leq t_k \leq N$, and the remaining number of available transmissions is denoted $1 \leq s_k \leq M$.

II. Source Process is Gaussian

A. The Solution in the n-D Case

The total estimation error over the N-step horizon can be expressed as

$$e^{\pi}_{(s,t)} = \sum_{k=0}^{N-1} \left\{ \left(b_k - \hat{b}_k\right)^T Q \left(b_k - \hat{b}_k\right) \right\} \tag{1}$$

where $Q \in \mathbb{R}^{n \times n}$ denotes a user-defined weighting matrix. The estimate $\hat{b}_k \in \mathbb{R}^n$ of b_k is defined as the following conditional expectation:

$$\hat{b}_k = E\left\{ b_k \mid (s_k, t_k); x_k \right\}$$

where $x_k \in \mathbb{R}^n$ denotes the observer output. The observer policy μ_k can be expressed as

$$\mu_k = \begin{cases} b_k & \text{if } b_k \in \mathcal{J}_{(s_k,t_k)} \\ \text{NT} & \text{if } b_k \in \mathcal{J}^c_{(s_k,t_k)} \end{cases}. \tag{2}$$

If the source data $b_k \in \mathcal{J}_{(s_k,t_k)}$, the observer transmits the data, and the estimator uses the transmitted data. If $b_k \in \mathcal{J}^c_{(s_k,t_k)}$, where $\mathcal{J}^c_{(s_k,t_k)}$ denotes the complement of $\mathcal{J}_{(s_k,t_k)}$, the observer does not transmit the data, but instead transmits a single bit datum indicating NT for no transmission. When the estimator receives the NT signal, it uses the expected value of the source data based on knowledge of the statistics of b_k. The observer policy utilizes the VOI of the source data to determine whether or not the data should be transmitted. Heuristically speaking, if the observer determines that the data falls within the region $\mathcal{J}^c_{(s_k,t_k)}$, the data is determined to have low informational value since it is close to the expected value of the source data. If the data falls within the set $\mathcal{J}_{(s_k,t_k)}$, it is far from the expected value, so it is determined to have high informational value. Data having low informational value is not transmitted by the observer; instead the NT datum is transmitted. Data having high informational value is transmitted, and the estimator simply uses the transmitted data. Since the estimator has knowledge of the statistics of the source data, it can minimize the overall error in the estimation of the source data by using the expected value of the data when the NT signal is received. In the following development, the procedure for calculating the optimal observation set $\mathcal{J}^*$ will be presented.

Based on (1), the DP equation can be used to express the optimal estimation error as [11]

$$\begin{aligned} e^*_{(s,t)} &= \min_{\mathcal{J}_{(s,t)}} \left\{ e^*_{(s-1,t-1)} - \left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)} \right) \times \right. \\ &\qquad \left. \int_{b \in \mathcal{J}^c} f(b)\, db + \int_{b \in \mathcal{J}^c} b^T Q b f(b)\, db \right\} \end{aligned} \tag{3}$$

given that b has zero mean, where the fact that $\int_{b \in \mathcal{J}} f(b)\, db = 1 - \int_{b \in \mathcal{J}^c} f(b)\, db$ was utilized.

B. Optimal Observation Set

In this section, the cost-to-go equation in (3) will be utilized to calculate an optimal observation set $\mathcal{J}^*$ within which the source data possesses high VOI. To that end, the optimal region $\mathcal{J}^{c*}$ will be calculated from (3) as the range of b that globally minimizes $e^*_{(s,t)}$. The set $\mathcal{J}^*$ will be used to develop an observer policy that only transmits data having high VOI.

The Gaussian PDF $f(b)$ in (3) can be expressed as

$$f(b) = \frac{1}{(2\pi)^{n/2} |P|^{1/2}} \exp\left\{ -\frac{1}{2} b^T P^{-1} b \right\} \tag{4}$$

To facilitate the following analysis, a linear transformation will be defined as

$$x = P^{-1/2} b \qquad b = P^{1/2} x. \tag{5}$$

Using the Jacobian determinant, (5) can be used to express the integration differential db as

$$db = \left| P^{1/2} \right| dx. \quad (6)$$

After using (5), the expression in (4) can be rewritten as

$$f(x) = \frac{1}{(2\pi)^{n/2} |P|^{1/2}} \exp\left\{ -\frac{1}{2} x^T x \right\} \quad (7)$$

where the fact that

$$\left[P^{-1/2} \right]^T = P^{-1/2} \quad (8)$$

was utilized. The motivation behind the linear transformation in (5) is based on the desire to facilitate the subsequent evaluation of the integrals in (3). After substituting (5) in (3), transforming into spherical coordinates, the optimal integration region $\mathcal{J}^{c*}$ minimizing the estimation error can be obtained as

$$b^{*T} P^{-1} b^* = \frac{n\left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)} \right)}{tr\,[QP]}. \quad (9)$$

where $tr\,[\cdot]$ denotes the trace of a matrix, and b^* denotes the optimal value of b.

III. Estimation Error Recursion

In this section, an analytical formulation for the estimation error recursion formula is derived. To derive an explicit mathematical expression for the optimal estimation error, the two integrals in (3) must be evaluated. Since the integration variable in this case is a vector, the expression will be transformed into spherical coordinates to facilitate the derivation.

After transforming the Gaussian PDF given in (7) to spherical coordinates, the PDF can be expressed as

$$f(r) = \left(\frac{1}{(2\pi)^{\frac{n}{2}} |P|^{1/2}} \right) \exp\left\{ -\frac{1}{2} r^2 \right\} \quad (10)$$

where $r \in \mathbb{R}$ denotes the radial distance in spherical coordinates.

A. Evaluating the Recursion Formula

By transforming the integral expression in (3) to spherical coordinates and utilizing the region defined in (9) to define the limits of integration, the optimal estimation error (cost to go) can be calculated as

$$\begin{aligned} e^*_{(s,t)} = {} & e^*_{(s-1,t-1)} - \left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)} \right) \left| P^{1/2} \right| \times \quad (11) \\ & \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{r^*} \left(f(r)\, r^{n-1} \sin^{n-2}\phi_1 \times \right. \\ & \left. \sin^{n-3}\phi_2 \cdots \sin\phi_{n-2} \right) dr d\phi_1 \cdots d\phi_{n-2} d\phi_{n-1} \\ & + \left| P^{1/2} \right| \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{r^*} \left(f(r)\, r^{n+1} g_{11} \cos^2\phi_1 \right. \\ & + \cdots + g_{nn} \sin^2\phi_1 \sin^2\phi_2 \cdots \sin^2\phi_{n-2} \times \\ & \left. \sin^2\phi_{n-1} \right) \sin^{n-2}\phi_1 \cdots \sin\phi_{n-2} \times \\ & dr d\phi_1 d\phi_2 \cdots d\phi_{n-2} d\phi_{n-1}. \end{aligned}$$

In (11), the limits of integration $r^* = \sqrt{\gamma}$ were calculated based on the transformed version of (9) in spherical coordinates, where $\gamma \in \mathbb{R}$ is defined as

$$\gamma = \frac{n\left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)} \right)}{tr\,[QP]}. \quad (12)$$

1) The Case of n Even: After performing the necessary integrations in (11), the optimal estimation error $e^*_{(s,t)}$ can be expressed for the case where n is even as follows:

$$\begin{aligned} e^*_{(s,t)} = {} & e^*_{(s-1,t-1)} - \left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)} \right) \times \quad (13) \\ & \left(\frac{(2\pi)^{\frac{n}{2}} c_{1e}(\gamma)}{2^{\frac{n}{2}-1} \left(\frac{n}{2} - 1 \right)!} \right) + tr\,(QP) \left(\frac{(2\pi)^{\frac{n}{2}} c_{2e}(\gamma)}{2^{\frac{n}{2}} \left(\frac{n}{2} \right)!} \right) \end{aligned}$$

where the auxiliary functions $c_{1e}(\gamma)$ and $c_{2e}(\gamma)$ are explicitly defined as

$$\begin{aligned} c_{1e}(\gamma) = {} & \left(\frac{1}{(2\pi)^{n/2} |P|^{1/2}} \right) \int_0^{\sqrt{\gamma}} r^{n-1} \exp\left\{ -\frac{1}{2} r^2 \right\} dr \\ = {} & \left(\frac{2^{\frac{n}{2}-1} \left(\frac{n}{2} - 1 \right)!}{(2\pi)^{\frac{n}{2}} |P|^{1/2}} \right) \times \quad (14) \\ & \left(1 - \left\{ \sum_{\substack{j=2 \\ j\ even}}^{n} \left\{ \frac{\left(\frac{1}{2} \right)^{\frac{j}{2}-1}}{\left(\frac{j}{2} - 1 \right)!} \right\} \gamma^{\frac{j}{2}-1} \right\} \right) \exp\left\{ -\frac{1}{2} \gamma \right\}. \end{aligned}$$

and

$$\begin{aligned} c_{2e}(\gamma) = {} & \left(\frac{1}{(2\pi)^{n/2} |P|^{1/2}} \right) \int_0^{\sqrt{\gamma}} r^{n+1} \exp\left\{ -\frac{1}{2} r^2 \right\} dr \quad (15) \\ = {} & \left(\frac{2^{\frac{n}{2}} \left(\frac{n}{2} \right)!}{(2\pi)^{\frac{n}{2}} |P|^{1/2}} \right) \times \\ & \left(1 - \left\{ \sum_{\substack{j=2 \\ j\ even}}^{n+2} \left\{ \frac{\left(\frac{1}{2} \right)^{\frac{j}{2}-1}}{\left(\frac{j}{2} - 1 \right)!} \right\} \gamma^{\frac{j}{2}-1} \right\} \right) \exp\left\{ -\frac{1}{2} \gamma \right\}. \end{aligned}$$

2) The Case of n Odd: In a manner similar to the case where n is even, the integral expression in (11) can be evaluated as follows for the case where n is odd:

$$\begin{aligned} e^*_{(s,t)} = {} & e^*_{(s-1,t-1)} - \sqrt{\frac{2}{\pi}} \left[\left(e^*_{(s-1,t-1)} - e^*_{(s,t-1)} \right) \times \right. \quad (16) \\ & \left(\frac{2^{\frac{n-3}{2}} \left(\frac{n-3}{2} \right)! \,(2\pi)^{\frac{n}{2}}}{(n-2)!} \right) c_{1o}(\gamma) \\ & \left. + tr\,(QP) \left(\frac{(2\pi)^{\frac{n}{2}}\, 2^{\frac{n-1}{2}} \left(\frac{n-1}{2} \right)!}{n!} \right) c_{2o}(\gamma) \right] \end{aligned}$$

where $c_{1o}(\gamma)$ and $c_{2o}(\gamma)$ are explicitly defined as

$$\begin{aligned} c_{1o}(\gamma) &= \int_0^{\sqrt{\gamma}} r^{n-1} f(r)\, dr \qquad (17) \\ &= \left(\frac{(n-2)!}{(2\pi)^{\frac{n}{2}} 2^{\frac{n-3}{2}} \left(\frac{n-3}{2}\right)! \, |P|^{1/2}} \right) \times \\ & \Bigg(\sqrt{2\pi} \left[\Phi\left(\sqrt{\gamma}\right) - \Phi(0) \right] \\ & - \left\{ \sum_{\substack{j=3 \\ j\ odd}}^{n} \left\{ \frac{2^{\frac{j-3}{2}} \left(\frac{j-3}{2}\right)!}{(j-2)!} \right\} \gamma^{\frac{j-2}{2}} \right\} \exp\left\{ -\frac{1}{2}\gamma \right\} \Bigg). \end{aligned}$$

and

$$\begin{aligned} c_{2o}(\gamma) &= \int_0^{\sqrt{\gamma}} r^{n+1} f(r)\, dr \qquad (18) \\ &= \left(\frac{n!}{2^{\frac{n-1}{2}} \left(\frac{n-1}{2}\right)! \, (2\pi)^{\frac{n}{2}} |P|^{1/2}} \right) \times \\ & \Bigg(\sqrt{2\pi} \left[\Phi\left(\sqrt{\gamma}\right) - \Phi(0) \right] \\ & - \left\{ \sum_{\substack{j=3 \\ j\ odd}}^{n+2} \left\{ \frac{2^{\frac{j-3}{2}} \left(\frac{j-3}{2}\right)!}{(j-2)!} \right\} \gamma^{\frac{j-2}{2}} \right\} \exp\left\{ -\frac{1}{2}\gamma \right\} \Bigg) \end{aligned}$$

respectively. In (17) and (18), $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) of the standard Gaussian random variable with zero-mean and unit variance, which is defined as

$$\Phi(b) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} \exp\left\{ -\frac{1}{2}\zeta^2 \right\} d\zeta \qquad (19)$$

for any $\zeta \in \mathbb{R}$.

IV. Source Process is Gauss-Markov

In this section, the optimal estimation technique outlined in the previous section will be applied to a system for which the source data is generated via a *Gauss-Markov process*.

In the case where the source process is Markov driven by an I.I.D. Gaussian process $\{w_k\}$ with zero mean, the source data is generated by the following model:

$$b_k = A b_{k-1} + w_{k-1} \qquad (20)$$

where b_k, $w_k \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$. The current value of the state b_k depends only on the value of the state at the previous time step (i.e., Markov property). If r_k time steps have passed since the last transmission was received, then the current value of the state b_k will depend on the value of the state r_k time steps in the past, i.e., b_{k-r}. If data were transmitted in the previous time step, then $r_k = 1$, and the current value b_k of the source data is given explicitly by (20). If $r_k > 1$, a linear regression can be performed on (20) to determine the current value of the state b_k in terms of b_{k-r} as

$$\begin{aligned} b_k &= A^r b_{k-r} + A^{r-1} w_{k-r} + A^{r-2} w_{k-r+1} \qquad (21) \\ & \quad + \cdots + A^1 w_{k-2} + A^0 w_{k-1}. \end{aligned}$$

To simplify the notation in the following analysis, let r denote the number of time units passed since the last transmission of a source output at time step k. So in the presence of noise $\{w_k\}$, the estimation error will increase with the number of time steps passed since the use of the channel for transmission. Based on (21), the expectation $E\left[b_k \mid b_{(k-r)}\right]$ (i.e., the expected value of b_k after r missed transmissions) can be expressed as

$$E\left[b_k \mid b_{(k-r)}\right] = A^r b_{k-r} \qquad (22)$$

where the fact that w_k has zero mean was utilized. Based on (22), it is apparent that the mean value of b_k varies with the number of missed transmissions r. Thus, the distribution of b can be expressed as

$$b \sim \mathcal{N}\left(A^r b_{k-r}, P_r\right) \qquad (23)$$

where $P_r \in \mathbb{R}^{n\times n}$ denotes the covariance of b_k after r missed transmissions.

A. Covariance Matrix Calculation

For the source process given in (20), the expected value of b and the covariance matrix will change with the number of missed transmissions r. In this section, the general formula for the covariance matrix based on r missed transmissions will be derived.

The covariance matrix P_r is defined as

$$P_r \triangleq E\left[(b - E[b])\, (b - E[b])^T \right]. \qquad (24)$$

In the n dimensional case, P_r can be expressed in matrix form as

$$P_r = \begin{bmatrix} \sigma_{1r}^2 & \sigma_{1r}\sigma_{2r} & \cdots & \sigma_{1r}\sigma_{nr} \\ \sigma_{1r}\sigma_{2r} & \sigma_{2r}^2 & \cdots & \sigma_{2r}\sigma_{nr} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1r}\sigma_{nr} & \cdots & \sigma_{(n-1)r}\sigma_{nr} & \sigma_{nr}^2 \end{bmatrix} \qquad (25)$$

where the elements of P_r (i.e., the variances σ_{ir}^2 and the covariances $\sigma_{ir}\sigma_{jr}$) in (25) can be obtained by performing the necessary multiplications in the expansion of (24) as follows:

$$\sigma_{ir}^2 = \sum_{q=1}^{r} \left(\sum_{m=1}^{n} A_{i,m}^{q-1} \sigma_m \right)^2 \qquad (26)$$

and the covariances can be expressed as

$$\sigma_{ir}\sigma_{jr} = \sum_{q=1}^{r} \left(\sum_{m=1}^{n} A_{i,m}^{q-1} \sigma_m \sum_{p=1}^{n} A_{j,p}^{q-1} \sigma_p \right) \qquad (27)$$

for $i, j = 1, \ldots, n$, where the fact that the noise w is I.I.D. was utilized.

Note that if $r = 1$, the variances and covariances reduce to σ_i^2 and $\sigma_i\sigma_j$, for $i, j = 1, \ldots, n$, respectively. In the scalar case, $n = 1$, and (26) can be used to show that the variance after r steps without transmissions can be calculated as [2]

$$\sigma_r^2 = \left(\sum_{k=1}^{r} A^{2(k-1)} \right) \sigma_b^2 \qquad (28)$$

where $A \in \mathbb{R}$, σ_r^2 denotes the variance of b after r time steps of no transmissions, and σ_b^2 denotes the variance of b after only a single time step without a transmission (i.e., when $r = 1$).

B. Optimal Observation Set

For the case where the source data is generated via the Gauss-Markov process in (20) and (21), the encoder and decoder policies can be derived in a manner very similar to the case where the source data is a Gaussian random variable. The main difference in the Gauss-Markov case is that the mean and covariance of the vector b can change at each time step based on the number of missed transmissions r, as expressed in (22), and (25) - (28).

The PDF in the Gauss-Markov case can be expressed as

$$f(b) = \left(\frac{1}{(2\pi)^{\frac{n}{2}} |P_r|^{1/2}} \right) \exp\left\{ -\frac{1}{2} (b-\mu)^T P_r^{-1} (b-\mu) \right\} \tag{29}$$

where $\mu = A^r b_{k-r}$, and P_r is defined in (25), (26), and (27). To derive the optimal observation set $\mathcal{J}_{(r_k,s_k,t_k)}$, the optimal cost to go is formulated as

$$\begin{aligned} e^*_{(r,s,t)} = \min_{\mathcal{J}_{(s,t)}} \Big\{ & e^*_{(1,s-1,t-1)} \\ & - \left(e^*_{(1,s-1,t-1)} - e^*_{(r+1,s-1,t-1)} \right) \int_{b \in \mathcal{J}^c} f(b)\, db \\ & + \int_{b \in \mathcal{J}^c} (b-\mu)^T Q (b-\mu) f(b)\, db \Big\}. \end{aligned} \tag{30}$$

The notation $e^*_{(r,s,t)}$ is used in this case to represent the optimal estimation error, since the estimation error depends on the three parameters: r, s, and t in the Gauss-Markov case. After transforming variables and performing the required integrations in (30), the optimal integration region can be expressed as

$$(b^* - A^r b_{k-r})^T P^{-1} (b^* - A^r b_{k-r}) = \gamma \tag{31}$$

where γ is defined as

$$\gamma \triangleq \frac{n \left(e^*_{(1,s-1,t-1)} - e^*_{(r+1,s,t-1)} \right)}{tr\left[Q P_r \right]}. \tag{32}$$

Hence, as in the case where the source data is purely Gaussian, the region minimizing the estimation error in the Gauss-Markov case is defined by an n dimensional *hyper-ellipsoid*. Unlike the Gaussian case, however, the center (i.e., μ) and shape (i.e., P_r) of the hyper-ellipsoidal region vary with the number of missed transmissions r.

C. Estimation Error Recursion

In a manner similar to the Gaussian source data case, the expression given in (30) can be integrated to express the optimal estimation error. Thus, following a procedure identical to that given in the previous section, expressions for the estimation error recursion formula for the Gauss-Markov case can be obtained for the cases where n is odd and where n is even. As before, the estimation error recursion can be evaluated explicitly for the case where n is even; and for the case where n is odd, the estimation error recursion contains the CDF $\Phi(\cdot)$ defined in (19). The explicit expressions for the estimation error recursions have been omitted here for brevity.

V. Simulation Results

Numerical simulations were created to test the performance of the proposed optimal estimation technique for the cases where the source data is generated via a purely Gaussian random process and via a Gauss-Markov process. For each simulation, a lookup table containing the optimal cost to go at each instant (s,t) was generated offline. The lookup table is used along with equations (13) or (16) for the purely Gaussian case, or with the corresponding equations for the Gauss-Markov case to calculate the optimal cost to go at each time step. For clarity of presentation, the simulation results presented in this paper were obtained using 2-D source data; however, the 2-D case effectively serves to illustrate the capability of this estimation technique to estimate incomplete multidimensional data. It is a trivial task to extend the 2-D results to n-D.

A. Source Process is Gaussian

For the purely Gaussian simulation, the source data is generated via a zero mean standard Gaussian random process with PDF and covariance matrix defined as

$$f(b) = \frac{1}{2\pi |P|^{1/2}} \exp\left\{ -b^T P^{-1} b \right\}, \; P = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}. \tag{33}$$

The initial conditions used in the simulation are

$$e^*_{(t,t)} = 0, \qquad e^*_{(0,t)} = nt, \qquad \forall\, t > 0.$$

Fig. 2 summarizes the results of the numerical simulation for the purely Gaussian source data case. In each of the six plot windows in Fig. 2, the point indicating the current value of the 2-D vector b_k is denoted as an 'o' or an '$\times$', where 'o' indicates that the data will be transmitted, and '$\times$' indicates that the data will not be transmitted. Toward the top right corner of each of the six plots in Fig. 2, the $[s_k, t_k]$ value indicates the value of s and t before the decision to transmit or not is made at that time step. Fig. 2 shows the optimal 2-D region J^{c*} at each time step for the case where $N = 6$ and $M = 4$.

B. Source Process is Gauss-Markov

For the 2-D Gauss-Markov case, the source data is generated via (20), where the process matrix $A \in \mathbb{R}^{2\times 2}$ is defined as

$$A \triangleq \begin{bmatrix} 0.5 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}$$

and $\{w_k\}$ is an I.I.D. Gaussian process with zero mean. The r-dependent covariance matrix and the PDF of $b \in \mathbb{R}^2$ can be expressed as given in (25) and (29), where $n = 2$. The elements of the covariance matrix are defined as

$$\sigma_{ir}\sigma_{jr} = \sum_{q=1}^{r} \left(\sum_{m=1}^{2} A^{q-1}_{i,m} \sigma_m \sum_{p=1}^{2} A^{q-1}_{j,p} \sigma_p \right) \tag{34}$$

for $i,j = 1,2$, where σ_m for $m = 1,2$ are elements of the covariance matrix for $r = 1$. The initial conditions used in the

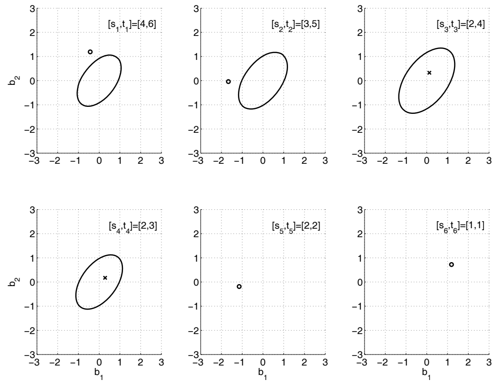

Fig. 2. The region minimizing the estimation error (the ellipse) and the source data point b_k at each time step for the case where $(M, N) = (4, 6)$.

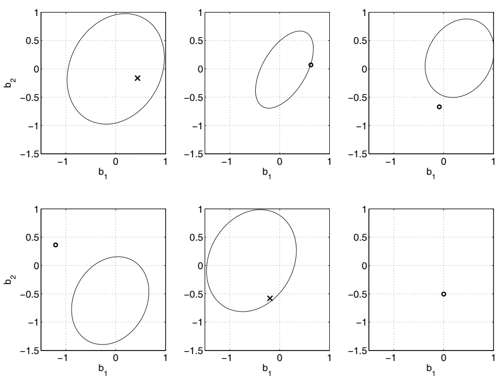

Fig. 3. The region minimizing the estimation error (the ellipse) and the source data point b_k ('o' or '$\times$') at each time step for the case Gauss-Markov case with $(M, N) = (4, 6)$.

Gauss-Markov simulation are

$$
\begin{aligned}
e^*_{(r,t,t)} &= 0, \\
e^*_{(r,0,t)} &= \sum_{m=r}^{r+t-1} \sum_{q=1}^{m} \left(\sum_{k=1}^{2} A^{q-1}_{i,k} \sigma_k \sum_{p=1}^{2} A^{q-1}_{j,p} \sigma_p \right)
\end{aligned}
$$

for $i, j = 1, 2, \forall\ t > 0$.

Fig. 3 shows the results from the simulation in the Gauss-Markov case. The dependence on the mean and covariance on the number r of missed transmissions results in the center and shape of the ellipses shown in Fig. 3 to vary with the time step k. Specifically, the means at each time step $1 \le k \le 6$ are (see (22)) $\mu_k = \begin{bmatrix} 0 & 0.10 & 0.32 & -0.11 & -0.57 & -0.27 \\ 0 & -0.01 & 0.19 & -0.62 & 0.09 & -0.04 \end{bmatrix}$, and the covariance matrices for $r = 1, 2$ are $P_1 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$ and $P_2 = \begin{bmatrix} 0.30 & 0.19 \\ 0.19 & 0.41 \end{bmatrix}$, respectively. The value of r is 2 for time steps $k = 2$ and $k = 6$ since there were no transmissions for $k = 1$ or $k = 5$.

VI. Conclusion

Optimal encoder policies are developed for a sensor that takes multi-dimensional measurements of some true state process. The criterion function to be minimized consists of a weighted cumulative quadratic estimation error under the constraint that over a time-window of length N, only $M < N$ measurements can be transmitted to a remote estimator. We show how the optimal policy is to transmit a measurement only if it lies outside of a certain hyper-ellipsoid which depends on the number of remaining channel-uses, the number of remaining time-steps, and the statistics of the measurement process. We show how dynamic programming can be used to find the optimal hyper-ellipsoid in a Gaussian scenario, as well as in a Gauss-Markov process, where the sensor is subject to Gaussian noise.

We also incorporate an arbitrary (positive-definite) weighting matrix that allows the user to specify which elements of the state vector are most valuable to estimate accurately. This is a first step in extending this research to address the larger issues of informational value and relevance when this problem is set in a closed-loop context (i.e., when the estimate is used by a control system to generate an input signal that alters the trajectory of the state that is being measured). In the previous (scalar-valued) work, there was no place for consideration of which dimensions of the sensor data might be most useful to control the system.

References

[1] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, MIT, 2000.

[2] O. C. Imer, "Optimal estimation and control under communication network constraints," Ph.D. Thesis, University of Illinois at Urbana-Champagne, 2005.

[3] Y. Q. Wang, H. Ye, and G. Z. Wang, "Fault detection of NCS based on eigendecomposition, adaptive evaluation and adaptive threshold," *Int'l Journal of Control*, vol. 80, no. 12, pp. 1903–1911, 2007.

[4] L. Meier, J. Peschon, and R. Dressler, "Optimal control of measurement subsystems," *IEEE Transactions on Automatic Control*, vol. 12, pp. 528–536, 1967.

[5] M. S. Andersland, "On the optimality of open-loop LQG measurement scheduling," *IEEE Transactions on Automatic Control*, vol. 40, no. 10, pp. 1796–1799, 1995.

[6] X. Feng and K. A. Loparo, "Active probing for information in control systems with quantized state measurements: A minimum entropy approach," *IEEE Trans. Automat. Contr.*, vol. 42, no. 2, pp. 216–238, 1997.

[7] H. Yue and H. Want, "Minimum entropy control of closed-loop tracking errors for dynamic stochastic systems," *IEEE Trans. Automat. Contr.*, vol. 48, pp. 118–122, 2003.

[8] N. Gans and J. W. Curtis, "A performance bound for decentralized moving horizon estimation," in *Dynamics of Information Systems: Theory and Applications*, Gainesville, FL, Jan. 2009.

[9] O. C. Imer and T. Basar, "Optimal estimation with limited measurements," *Int'l Journal of Systems, Control and Communications*, vol. 2, no. 1-2, pp. 5–29, 2010.

[10] G. M. Lipsa and N. C. Martins, "Certifying the optimality of a distributed state estimation system via majorization theory," Institute for Systems Research Technical Reports, Tech. Rep., 2009.

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