Distortion-transmission trade-off in real-time transmission of Markov sources

Jhelum Chakravorty
Electrical and Computer Engg., McGill University
Montreal, Quebec, H3A 0E9
Email: jhelum.chakravorty@mail.mcgill.ca

Aditya Mahajan
Electrical and Computer Engg., McGill University
Montreal, Quebec, H3A 0E9
Email: aditya.mahajan@mcgill.ca

***Abstract*—The problem of optimal real-time transmission of a Markov source under constraints on the expected number of transmissions is considered, both for the discounted and long term average cases. This setup is motivated by applications where transmission is sporadic and the cost of switching on the radio and transmitting is significantly more important than the size of the transmitted data packet. For this model, we characterize the distortion-transmission function, i.e., the minimum expected distortion that can be achieved when the expected number of transmissions is less than or equal to a particular value. In particular, we show that the distortion-transmission function is a piecewise linear, convex, and decreasing function. We also give an explicit characterization of each vertex of the piecewise linear function. The results are illustrated using an example of a birth-death Markov chain.**

I. Introduction

In many applications of *real-time* communication systems, the transmitter is a battery powered device that transmits over a wireless packet-switched network; the cost of switching on the radio and transmitting a packet is significantly more important than the size of the data packet. Therefore, the transmitter does not transmit all the time; but when it does transmit, the transmitted packet is as big as needed to communicate the current source realization. In this paper, we characterize a fundamental trade-off between the real-time (i.e. zero-delay) distortion and the average number of transmissions in such systems.

In particular, we consider a transmitter that observes a first-order Markov source. At each time instant, based on the current source symbol and the history of its past decisions, the transmitter determines whether or not to transmit the current source symbol. If the transmitter does not transmit, the receiver must estimate the source symbol using the previously transmitted values. A per-step distortion function measures the fidelity of estimation. We are interested in characterizing the optimal transmission and estimation strategies that minimize the expected distortion over an infinite horizon under a constraint on the expected number of transmissions.

The communication system described above is much simpler than the general real-time communication setup due to the following feature: whenever the transmitter transmits, it sends the current realization of the source to the receiver. These transmitted events *reset* the system. In addition, we impose certain symmetry assumptions on the model, which ensure that there is a single reset state. Exploiting these special features we show that threshold-based strategies are optimal at the transmitter; the optimal transmission strategy randomizes between two threshold-based strategies; the randomization takes place only at one state.

Several variations of the communication system described above have been considered in the literature. The most closely related models are [1]–[5] (where transmission is assumed to be unconstrained but expensive). Other related work includes censoring sensors [6], [7] (where a sensor takes a measurement and decides whether to transmit it or not; in the context of sequential hypothesis testing), estimation with measurement cost [8], [9] (where the receiver decides when the sensor should transmit), sensor sleep scheduling [10], [11] (where the sensor is allowed to sleep for a pre-specified amount of time); and event-based communication [12], [13] (where the sensor transmits when a certain event takes place).

Throughout this paper, we use the following notation. $\mathbb{Z}$, $\mathbb{Z}_{\ge 0}$ and $\mathbb{Z}_{>0}$ denote the set of integers, the set of non-negative integers and the set of strictly positive integers respectively. Upper-case letters (e.g., X, Y) denote random variables; corresponding lower-case letters (e.g. x, y) denote their realizations. $X_{1:t}$ is a short hand notation for the vector $(X_1, \dots, X_t)$. Given a matrix A, A_{ij} denotes its (i, j)-th element, A_i denotes its i-th row, $A^\intercal$ denotes its transpose. We index the matrices by sets of the form $\{-k, \dots, k\}$; so the indices take both positive and negative values. I_k denotes the identity matrix of dimension $k \times k$, $k \in \mathbb{Z}_{>0}$. $\mathbf{1}_k$ denotes $k \times 1$ vector of ones. $\langle v, w \rangle$ denotes the inner product between vectors v and w, $\mathbb{E}[\cdot]$ denotes the expectation of a random variable, and $\mathbb{1}\{\cdot\}$ denotes the indicator function of a statement. We follow the convention of calling a sequence $\{a_k\}_{k=0}^{\infty}$ increasing when $a_1 \le a_2 \le \cdots$. If all the inequalities are strict, then we call the sequence strictly increasing.

II. Problem formulation

A. The communication system

In this paper, we investigate the following communication setup. A transmitter causally observes a first-order Markov source $\{X_t\}_{t=0}^{\infty}$, where $X_t \in \mathbb{Z}$ and the initial state $X_0 = 0$. At each time, it may choose whether or not to transmit the current source observation. This decision is denoted by $U_t \in \{0, 1\}$, where $U_t = 0$ denotes no transmission and $U_1 = 1$ denotes transmission. The decision to transmit is made using a *transmission strategy* $f = \{f_t\}_{t=0}^{\infty}$, where

$$U_t = f_t(X_{0:t}, U_{0:t-1}). \tag{1}$$

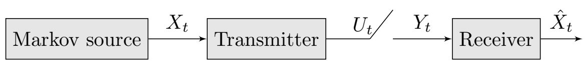


Fig. 1: A block diagram depicting the communication system considered in this paper.

We use the short-hand notation $X_{0:t}$ to denote the sequence $(X_0, \dots, X_t)$. Similar interpretations hold for $U_{0:t-1}$.

The transmitted symbol, which is denoted by Y_t, is given by

$$Y_t = \begin{cases} X_t, & \text{if } U_t = 1; \\ \mathfrak{E}, & \text{if } U_t = 0, \end{cases}$$

where $Y_t = \mathfrak{E}$ denotes no transmission.

The receiver causally observes $\{Y_t\}_{t=0}^{\infty}$ and generates a source reconstruction $\{\hat{X}_t\}_{t=0}^{\infty}$ (where $\hat{X}_t \in \mathbb{Z}$) in real-time using an *estimation strategy* $g = \{g_t\}_{t=0}^{\infty}$, i.e.,

$$\hat{X}_t = g_t(Y_{0:t}). \tag{2}$$

The fidelity of the reconstruction is measured by a per-step distortion $d(X_t - \hat{X}_t)$, where $d \colon \mathbb{Z} \to \mathbb{R}_{\geq 0}$.

Fig. 1 shows a communication system as described above. We impose the following assumptions on the model.

- **(A1)** The transition matrix P of the Markov source is a Toeplitz matrix with decaying off-diagonal terms, i.e., $P_{ij} = p_{|i-j|}$, where $\{p_n\}_{n=0}^{\infty}$ is a decreasing non-negative sequence and $p_1 > 0$.
- **(A2)** The distortion function is even and increasing on $\mathbb{Z}_{\geq 0}$, i.e., for all $e \in \mathbb{Z}_{\geq 0}$, $d(e) = d(-e)$ and $d(e) \leq d(e+1)$. Furthermore, $d(0) = 0$ and $d(e) \neq 0, \forall e \neq 0$.

An example of a source and a distortion function that satisfy the above assumptions is the following:

Example 1 Consider an aperiodic, symmetric, birth-death Markov chain defined over $\mathbb{Z}$ with the transition probability matrix as given by the following:

$$P_{ij} = \begin{cases} p, & \text{if } |i-j| = 1; \\ 1-2p, & \text{if } i = j; \\ 0, & \text{otherwise,} \end{cases}$$

where we assume that $p \in (0, \frac{1}{2})$. Let the distortion function be $d(e) = |e|$. P satisfies (A1) and $d(e)$ satisfies (A2). □

B. The optimization problem

The objective is to choose the transmission and estimation strategies (called the *communication strategy* in short) to minimize the expected distortion under a constraint on the expected number of transmissions. We investigate two variations of this objective: the discounted setup and the long-term average setup. In order to do that, let us first define the performance measures corresponding to a transmission and estimation strategy (f, g), denoted by $D_\beta(f, g)$ and $N_\beta(f, g)$, $\beta \in (0, 1]$ as follows:

1) Performance measures for the discounted setup: Given a communication strategy (f, g) and a discount factor $\beta \in (0, 1)$, let

$$D_\beta(f, g) \coloneqq (1-\beta)\mathbb{E}^{(f,g)}\Big[\sum_{t=0}^{\infty} \beta^t d(X_t - \hat{X}_t) \;\Big|\; X_0 = 0\Big]$$

denote the expected discounted distortion and

$$N_\beta(f, g) \coloneqq (1-\beta)\mathbb{E}^{(f,g)}\Big[\sum_{t=0}^{\infty} \beta^t U_t \;\Big|\; X_0 = 0\Big]$$

denote the expected discounted number of transmissions.

2) Performance measures for the long-term average setup: The long-term average setup is similar. Given a communication strategy (f, g), let

$$D_1(f, g) \coloneqq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{(f,g)}\Big[\sum_{t=0}^{T-1} d(X_t - \hat{X}_t) \;\Big|\; X_0 = 0\Big]$$

denote the expected long-term average distortion and

$$N_1(f, g) \coloneqq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{(f,g)}\Big[\sum_{t=0}^{T-1} U_t \;\Big|\; X_0 = 0\Big]$$

denote the expected long-term average number of transmissions.

We now state the constrained optimization problem as follows:

Problem *Given* $\alpha \in (0, 1)$ *and* $\beta \in (0, 1]$, *find a strategy* (f^*, g^*) *such that,*

$$D^*_\beta(\alpha) \coloneqq D_\beta(f^*, g^*) \coloneqq \inf_{(f,g): N_\beta(f,g) \leq \alpha} D_\beta(f, g) \tag{CON}$$

where the infimum is taken over all history-dependent communication strategies of the form (1) *and* (2).

C. The main result

The function $D^*_\beta(\alpha)$, $\beta \in (0, 1]$, represents the minimum expected distortion that can be achieved when the expected number of transmissions are less than or equal to α. It is analogous to the distortion-rate function in classical Information Theory; for that reason, we call it the *distortion-transmission function*.

By definition, $D^*_\beta(\alpha)$ is convex and decreasing in α. In this paper, we characterize the shape of $D^*_\beta(\alpha)$ for a class of Markov sources and distortion functions (those that satisfy (A1) and (A2)). In particular, we show that $D^*_\beta(\alpha)$ is piecewise linear (in addition to being convex and decreasing). We derive closed form expressions for each vertices; thus, completely characterizing the curve.

Specifically, we show that each point on the distortion-transmission function (i.e. the optimal distortion for a given value of α) is achieved by a communication strategy that is of the following form:

- Let Z_t be the most recently transmitted symbol up to time t. Then, the optimal estimation strategy is

$$g^*(Y_{0:t}) = Z_t. \tag{3}$$

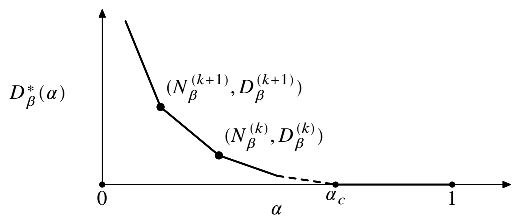


Fig. 2: The distortion-transmission function $D^*_\beta(\alpha)$ for a symmetric Markov source and even and increasing distortion function. $D^*_\beta(\alpha)$ is piecewise linear, convex, and decreasing.

- Let $E_t = X_t - Z_{t-1}$ and $f^{(k)}$ be a threshold-based strategy given by

$$f^{(k)}(X_t, Y_{0:t-1}) = \begin{cases} 1, & \text{if } |E_t| \geq k; \\ 0, & \text{if } |E_t| < k. \end{cases} \tag{4}$$

  Then, the optimal transmission strategy is a possibly randomized strategy that, at each stage, picks $f^{(k^*)}$ with probability θ^* and picks $f^{(k^*+1)}$ with probability $(1-\theta^*)$; where k^* is the largest k such that $N_\beta(f^{(k)}, g^*) \geq \alpha$ and θ^* is chosen such that

$$\theta^* N_\beta(f^{(k^*)}, g^*) + (1-\theta^*) N_\beta(f^{(k^*+1)}, g^*) = \alpha.$$

Note that this randomized strategy can also be written as

$$f^*(e) = \begin{cases} 0, & \text{if } |e| < k^*; \\ 0, & \text{w.p. } 1-\theta^*, \text{ if } |e| = k^*; \\ 1, & \text{w.p. } \theta^*, \text{ if } |e| = k^*; \\ 1, & \text{if } |e| > k^*. \end{cases} \tag{5}$$

Thus, it randomizes only at states $\{-k^*, k^*\}$.

The corresponding distortion-transmission function is a piecewise-linear function with vertices given by $(N_\beta^{(k)}, D_\beta^{(k)})$, where

$$D_\beta^{(k)} = D_\beta(f^{(k)}, g^*) \quad \text{and} \quad N_\beta^{(k)} = N_\beta(f^{(k)}, g^*).$$

In addition, $D_\beta^{(1)} = 0$. Therefore,

$$D^*_\beta(\alpha) = 0, \quad \forall \alpha > \alpha_c := N_\beta^{(1)} = \beta(1-p_0).$$

We show that $\{N_\beta^{(k)}\}_{k=0}^\infty$ is a decreasing sequence and $\{D_\beta^{(k)}\}_{k=0}^\infty$ is an increasing sequence. Consequently, the distortion-transmission function is convex and decreasing. See Fig. 2 for an illustration.

III. Proof of the main result

A. Lagrange relaxations

The Lagrange relaxation of Problem (CON) is the following: for any $\beta \in (0,1]$ and $\lambda \geq 0$, find a strategy (f^*, g^*) such that

$$C^*_\beta(\lambda) := C_\beta(f^*, g^*; \lambda) := \inf_{(f,g)} C_\beta(f, g; \lambda) \tag{LAG}$$

where $C_\beta(f,g;\lambda) = D_\beta(f,g) + \lambda N_\beta(f,g)$ and the infimum is taken over all history-dependent communication strategies of the form (1) and (2).

It is shown in [14] and [15] for discounted setup and long-term average setup respectively, that for problem (LAG) the optimal estimation strategy is time homogeneous, Kalman-like and is independent of the transmission startegy (as given by (3)). The optimal transmission strategy is of threshold-type and the optimal thresholds are time homogeneous (as given by (4)).

Furthermore, the optimal thresholds are characterized as follows. Let $\lambda_\beta^{(k)}$ be the value of the Lagrange multiplier for which one is indifferent between strategies $f^{(k)}$ and $f^{(k+1)}$ when starting from state 0, i.e., $\lambda_\beta^{(k)}$ is such that

$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)}), \tag{6}$$

where $C_\beta^{(k)}(\lambda) = D_\beta^{(k)} + \lambda N_\beta^{(k)}$ denotes the Lagrange performance of strategy $(f^{(k)}, g^*)$. Such a sequence of $\{\lambda_\beta^{(k)}\}_{k=0}^\infty$ can be computed based on $D_\beta^{(k)}$ and $N_\beta^{(k)}$ as follows:

Proposition 1 *For any $\beta \in (0,1]$, the sequence $\{\lambda_\beta^{(k)}\}_{k=0}^\infty$ given by*

$$\lambda_\beta^{(k)} := \frac{D_\beta^{(k+1)} - D_\beta^{(k)}}{N_\beta^{(k)} - N_\beta^{(k+1)}} \tag{7}$$

satisfies (6) *for all $k \in \mathbb{Z}_{\geq 0}$. Under (A2), $\lambda_\beta^{(k)} > 0$ for all $k \in \mathbb{Z}_{\geq 0}$.* □

Proposition 1 is a strainght forward consequence of [14, Lemma 7] and [15, Eq. (20)]. Also, please see [16] for the proof.

B. The constrained optimization problems

We now analyze the constrained optimization problem (CON). For that matter, define Bernoulli randomized strategy and Bernoulli randomized simple strategy.

Definition 1 Suppose we are given two (non-randomized) time-homogeneous strategies f_1 and f_2 and a randomization parameter $\theta \in (0,1)$. The *Bernoulli randomized strategy* (f_1, f_2, θ) is a strategy that randomizes between f_1 and f_2 at each stage; choosing f_1 with probability θ and f_2 with probability $(1-\theta)$. Such a strategy is called a Bernoulli randomized *simple* strategy if f_1 and f_2 differ on exactly one state i.e. there exists a state e_0 such that

$$f_1(e) = f_2(e), \quad \forall e \neq e_0.$$ □

Define

$$k^*_\beta(\alpha) = \sup\{k \in \mathbb{Z}_{\geq 0} : N_\beta(f^{(k)}, g^*) \geq \alpha\} \tag{8}$$

and

$$\theta^*_\beta(\alpha) = \frac{\alpha - N_\beta(f^{(k^*_\beta(\alpha)+1)}, g^*)}{N_\beta(f^{(k^*_\beta(\alpha))}, g^*) - N_\beta(f^{(k^*_\beta(\alpha)+1)}, g^*)}. \tag{9}$$

For ease of notation, we use $k^* = k^*_\beta(\alpha)$ and $\theta^* = \theta^*_\beta(\alpha)$. By definition, $\theta^* \in [0,1]$ and

$$\theta^* N_\beta(f^{(k^*)}, g^*) + (1-\theta^*) N_\beta(f^{(k^*+1)}, g^*) = \alpha. \tag{10}$$

Theorem 1 *Let f^* be the Bernoulli randomized simple strategy $(f^{(k^*)}, f^{(k^*+1)}, \theta^*)$. of the form* (5)*. Then (f^*, g^*) is optimal for the constrained Problem* (CON) *when $\beta \in (0,1]$.* □

TABLE I: Values of $D_\beta^{(k)}$, $N_\beta^{(k)}$ and $\lambda_\beta^{(k)}$ for different values of k and β for the birth-death Markov chain of Example 1 with $p = 0.3$.

(a) For $\beta = 0.9$

k	$D_\beta^{(k)}$	$N_\beta^{(k)}$	$\lambda_\beta^{(k)}$
0	0	1	0
1	0	0.5400	1.0989
2	0.4576	0.1236	4.1021
3	0.7695	0.0475	9.2839
4	1.0066	0.0220	16.2509
5	1.1844	0.0111	24.4478
6	1.3130	0.0058	33.4121
7	1.4029	0.0031	42.8289
8	1.4638	0.0017	52.5042
9	1.5040	0.0009	62.3245
10	1.5298	0.0005	72.2255

(b) For $\beta = 0.95$

k	$D_\beta^{(k)}$	$N_\beta^{(k)}$	$\lambda_\beta^{(k)}$
0	0	1	0
1	0	0.5700	1.1050
2	0.4790	0.1365	4.3657
3	0.8282	0.0565	10.6058
4	1.1218	0.0288	19.9550
5	1.3715	0.0163	32.0869
6	1.5811	0.0098	46.4727
7	1.7536	0.0061	62.5651
8	1.8927	0.0039	79.8921
9	2.0028	0.0025	98.0854
10	2.0884	0.0016	116.8739

(c) For $\beta = 1.0$

k	$D_\beta^{(k)}$	$N_\beta^{(k)}$	$\lambda_\beta^{(k)}$
0	0	1	0
1	0	0.6000	1.1111
2	0.5000	0.1500	4.6667
3	0.8889	0.0667	12.3810
4	1.2500	0.0375	25.9259
5	1.6000	0.0240	46.9697
6	1.9444	0.0167	77.1795
7	2.2857	0.0122	118.2222
8	2.6250	0.0094	171.7647
9	2.9630	0.0074	239.4737
10	3.0000	0.0060	323.0159

Proof: The proof relies on the following characterization of the optimal strategy stated in [17, Proposition 1.2]. The characterization was stated for the long-term average setup but a similar result can be shown for the discounted case as well, for example, by using the approach of [18]. Also, see [19, Theorem 8.1] for a similar sufficient condition for general constrained optimization problem.

A (possibly randomized) strategy (f°, g°) is optimal for a constrained optimization problem with $\beta \in (0, 1]$ if the following conditions hold:

(C1) $N_\beta(f^\circ, g^\circ) = \alpha$,

(C2) There exists a Lagrange multiplier $\lambda^\circ \geq 0$ such that (f°, g°) is optimal for $C_\beta(f, g; \lambda^\circ)$.

We will show that the strategies (f^*, g^*) satisfy (C1) and (C2) with $\lambda^\circ = \lambda_\beta^{(k^*)}$.

The strategy (f^*, g^*) satisfies (C1) due to (10). For $\lambda = \lambda_\beta^{(k^*)}$, both $f^{(k^*)}$ and $f^{(k^*+1)}$ are optimal for $C_\beta(f, g; \lambda)$. Hence, any strategy randomizing between them, in particular f^*, is also optimal for $C_\beta(f, g; \lambda)$. Hence (f^*, g^*) satisfies (C2). Therefore, by [17, Proposition 1.2], (f^*, g^*) is optimal for Problem (CON). ■

Theorem 2 *The distortion-transmission function is given by*

$$D_\beta^*(\alpha) = \theta^* D_\beta(f^{(k^*)}, g^*) + (1 - \theta^*) D_\beta(f^{(k^*+1)}, g^*). \quad (11)$$

Furthermore, $D_\beta^*(\alpha)$ is a continuous, piecewise linear, decreasing, and convex function of α. □

Proof: The form of $D_\beta^*(\alpha)$ given in (11) follows immediately from the fact that (f^*, g^*) is a Bernoulli randomized simple strategy. By definition, $D_\beta^*(\alpha)$ is decreasing and convex in α.

To show piecewise linearity, define for any $k \in \mathbb{Z}_{\geq 0}$, $\alpha^{(k)} = N_\beta(f^{(k)}, g^*)$, and consider any $\alpha \in (\alpha^{(k+1)}, \alpha^{(k)})$. Then,

$$k_\beta^*(\alpha^{(k)}) = k, \quad \text{and} \quad \theta_\beta^*(\alpha^{(k)}) = 1.$$

Hence $D_\beta^*(\alpha^{(k)}) = D_\beta(f^{(k)}, g^*)$. Thus, by (9), $\theta^* = (\alpha - \alpha^{(k+1)})/(\alpha^{(k)} - \alpha^{(k+1)})$, and by (11),

$$D_\beta^*(\alpha) = \theta^* D_\beta^*(\alpha^{(k)}) + (1 - \theta^*) D_\beta^*(\alpha^{(k+1)}).$$

Therefore $D_\beta^*(\alpha)$ is piecewise linear and continuous. ■

It follows from the argument given in the proof above that $\{(\alpha^{(k)}, D_\beta^*(\alpha^{(k)}))\}_{k=0}^\infty$ are the vertices of the piecewise linear function D_β^*. See Fig. 2 for an illustration.

IV. An example: Aperiodic, symmetric birth-death Markov chain

We state the main results for Example 1 here. The proofs, however, are not given here due to page limit. Please refer to [16] for the detailed proofs.

Lemma 1 1) *For $\beta \in (0, 1)$,*

$$D_\beta^{(k)} = \frac{\sinh(km_\beta) - k \sinh(m_\beta)}{2 \sinh^2(km_\beta/2) \sinh(m_\beta)};$$

$$N_\beta^{(k)} = \frac{2\beta p \sinh^2(m_\beta/2) \cosh(km_\beta)}{\sinh^2(km_\beta/2)} - (1 - \beta).$$

2) *For $\beta = 1$,*

$$D_1^{(k)} = \frac{k^2 - 1}{3k}; \quad N_1^{(k)} = \frac{2p}{k^2}, \textit{ and}$$

$$\lambda_1^{(k)} = \frac{k(k+1)(k^2 + k + 1)}{6p(2k+1)}.$$ □

When $p = 0.3$, the values of $D_\beta^{(k)}$, $N_\beta^{(k)}$, and $\lambda_\beta^{(k)}$ for different values of k and β are shown in Table I.

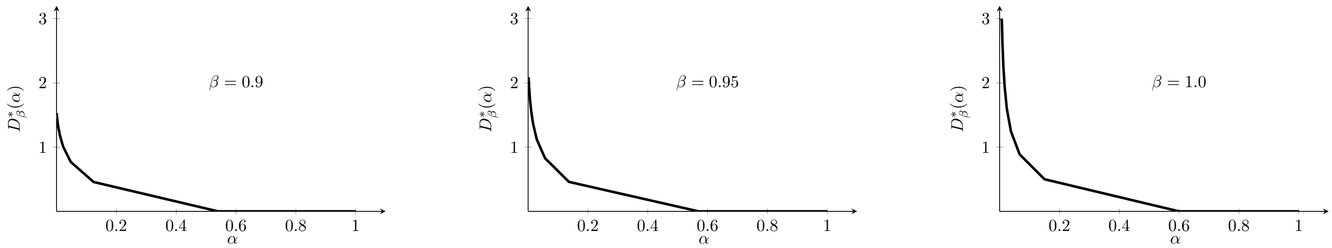
Lemma 2 1) *For $\beta \in (0, 1)$, k_β^* is given by the maximum k that satisfies the following inequality*

$$\frac{2\cosh(km_\beta)}{\cosh(km_\beta) - 1} \geq \frac{1 + \alpha - \beta}{\beta p(\cosh(m_\beta) - 1)}.$$

2) *For $\beta = 1$, k_1^* is given by the following equation*

$$k_1^* = \left\lfloor \sqrt{\frac{2p}{\alpha}} \right\rfloor.$$ □

Using the above results, we can plot the distortion-transmission function $D_\beta^*(\alpha)$. See Fig. 3 for the plot of $D_\beta^*(\alpha)$ vs α for different values of β (all for $p = 0.3$). An alternative way to plot this curve is to draw the vertices $(N_\beta^{(k)}, D_\beta^{(k)})$ using the data in Table I to compute the optimal (randomized) strategy for a particular value of α.



(a) $D^*_\beta(\alpha)$ vs α for $\beta = 0.9$

(b) $D^*_\beta(\alpha)$ vs α for $\beta = 0.95$

(c) $D^*_\beta(\alpha)$ vs α for $\beta = 1.0$

Fig. 3: Plots of $D^*_\beta(\alpha)$ vs α for different β for the birth-death Markov chain of Example 1 with $p = 0.3$.

As an example, suppose we want to identify the optimal strategy at $\alpha = 0.5$ for the birth-death Markov chain of Example 1 with $p = 0.3$ and $\beta = 0.9$. Recall that k^* is the largest value of k such that $N^{(k)}_\beta \le \alpha$. Thus, from (8), we get that $k^* = 1$. Then, by (9), $\theta^* = (\alpha - N^{(2)}_\beta)/(N^{(1)}_\beta - N^{(2)}_\beta) = 0.9039$. Let $f^* = (f^{(1)}, f^{(2)}, \theta^*)$. Then the Bernoulli randomized simple strategy (f^*, g^*) is optimal for Problem (CON) with $\beta = 0.9$. Furthermore, by (11), $D^*_\beta(\alpha) = 0.044$.

V. Conclusion

We characterized the distortion-transmission function for transmitting a first-order symmetric Markov source in real-time with constraints on the expected number of transmissions.

Our result depends critically on establishing the following structure of optimal communication strategies. There is no loss of optimality in restricting attention to threshold based transmission strategies and as long as the transmission strategy belongs to this class, the optimal estimation stratgey is independent of the choice of the threshold.

As a consequence of this structure, the optimal estimation strategy is known, and we only have to identify the optimal transmission strategy. We look at the Lagrange relaxation, compute the performance of an arbitrary threshold based transmission strategy, identify the set of Lagrange multipliers for which an arbitrary threshold based strategy is optimal, and then use these features to identify the optimal strategy for the constrained optimization problem.

References

[1] O. C. Imer and T. Basar, “Optimal estimation with limited measurements,” in *Proc. of Joint 44th IEEE Conference on Decision and Control and European Control Conference, Seville, Spain, December 12-15*, 2005, pp. 1029–1034.

[2] Y. Xu and J. P. Hespanha, “Optimal communication logics for networked control systems,” in *Proc. of the 43rd IEEE Conference on Decision and Control, The Bahamas, December 14-17*, 2004, pp. 3527–3532.

[3] G. M. Lipsa and N. Martins, “Remote state estimation with communication costs for first-order LTI systems,” *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2013–2025, Sep. 2011.

[4] A. Nayyar, T. Basar, D. Teneketzis, and V. Veeravalli, “Optimal strategies for communication and remote estimation with an energy harvesting sensor,” *IEEE Trans. Autom. Control*, vol. 58, no. 9, pp. 2246–2260, Sep. 2013.

[5] A. Molin and S. Hirche, “An iterative algorithm for optimal event-triggered estimation,” in *Proc. of 4th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS’12), TU Eindhoven, Netherlands, June 6-8*, 2012, pp. 64–69.

[6] C. Rago, P. Willett, and Y. Bar-Shalom, “Censoring sensors: A low-communication rate scheme for distributed detection,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 554–568, Apr. 1996.

[7] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, “Decentralized detection with censoring sensors,” *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1362–1373, Apr. 2008.

[8] M. Athans, “On the determination of optimal costly measurement strategies for linear stochastic systems,” *Automatica*, vol. 8, no. 4, pp. 397–412, Jul. 1972.

[9] J. Geromel, “Global optimization of measurement startegies for linear stochastic systems,” *Automatica*, vol. 25, no. 2, pp. 293–300, Mar. 1989.

[10] D. Shuman and M. Liu, “Optimal sleep scheduling for a wireless sensor network node,” in *Proc. of the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, October 29-November 1*, 2006, pp. 1337–1341.

[11] M. Sarkar and R. L. Cruz, “Analysis of power management for energy and delay trade-off in a WLAN,” in *Proc. of the Conference on Information Sciences and Systems, March*, 2004, pp. 1–6.

[12] K. J. Åström, *Analysis and Design of Nonlinear Control Systems*. Berlin, Heidelberg: Springer, 2008, ch. Event based control, pp. 127–147.

[13] X. Meng and T. Chen, “Optimal sampling and performance comparison of periodic and event based impulse control,” *IEEE Transactions of Automatic Control*, vol. 57, no. 12, pp. 3252–3259, Jan. 2012.

[14] J. Chakravorty and A. Mahajan, “On the optimal thresholds in remote state estimation with communication costs,” in *Proc. of 53rd IEEE Conference on Decision and Control, CDC 2014, Los Angeles, CA, USA, December 15-17*, 2014, pp. 1041–1046.

[15] ——, “Average cost optimal threshold strategies for remote estimation with communication cost,” in *Proc. of 52nd Allerton conference on communication, control, and computing, Illinois, USA, October 1-3*, 2014, pp. 335–342.

[16] ——, “Distortion-transmission trade-off in real-time transmission of Markov sources,” Dec 2014. [Online]. Available: http://arxiv.org/abs/1412.3199

[17] L. I. Sennott, “Computing average optimal constrained policies in stochastic dynamic programming,” *Probability in the Engineering and Informational Sciences*, vol. 15, pp. 103–133, Jan. 2001.

[18] V. Borkar, “A convex analytic approach to Markov decision processes,” *Probability Theory and Related Fields*, vol. 78, no. 4, pp. 583–602, Aug. 1988.

[19] D. Luenberger, *Optimization by Vector Space Methods*, ser. Professional Series. Wiley, 1968.