Les Cahiers du GERAD ISSN: 0711–2440

Fundamental limits of remote estimation of Markov processes under communication constraints

J. Chakravorty
A. Mahajan

G–2015–53

May 2015

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Fundamental limits of remote estimation of Markov processes under communication constraints

Jhelum Chakravorty
Aditya Mahajan

GERAD & Department of Electrical and Computer Engineering, McGill University, Montréal (Québec) Canada

jhelum.chakravorty@mail.mcgill.ca
aditya.mahajan@mcgill.ca

May 2015

Les Cahiers du GERAD
G–2015–53

Abstract: The fundamental limits of remote estimation of Markov processes under communication constraints are presented. The remote estimation system consists of a sensor and an estimator. The sensor observes a discrete-time Markov process, which is a symmetric countable state Markov source or a Gauss-Markov process. At each time, the sensor either transmits the current state of the Markov process or does not transmit at all. Communication is noiseless but costly. The estimator estimates the Markov process based on the transmitted observations. In such a system, there is a trade-off between communication cost and estimation accuracy. Two fundamental limits of this trade-off are characterized for infinite horizon discounted cost and average cost setups. First, when each transmission is costly, we characterize the minimum achievable cost of communication plus estimation error. Second, when there is a constraint on the average number of transmissions, we characterize the minimum achievable estimation error. Transmission and estimation strategies that achieve these fundamental limits are also identified.

Acknowledgments: This paper was presented in part in the Proceedings of the 52nd Annual Allerton Conference on Communication, Control, and Computing, 2014, the Proceedings of the 53rd Conference on Decision and Control, 2014, the Proceedings of the IEEE Information Theory Workshop, 2015 and the Proceedings of the IEEE International Symposium on Information Theory, 2015.

This work was supported in part by Fonds de recherche du Québec – Nature et technologies (FRQNT) Team Grant PR-173396.

1 Introduction

1.1 Motivation and literature overview

In many applications such as networked control systems, sensor and surveillance networks, and transportation networks, etc., data must be transmitted sequentially from one node to another under a strict delay deadline. In many of such *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 estimation error and the cost or average number of transmissions in such systems.

In particular, we consider a sensor that observes a first-order Markov process. At each time instant, based on the current source symbol and the history of its past decisions, the sensor determines whether or not to transmit the current state. If the sensor does not transmit, the receiver must estimate the state using the previously transmitted values. A per-step distortion function measures the estimation error. We investigate two fundamental trade-offs in this setup: (i) when there is a cost associated with each communication, what is the minimum expected estimation error and communication cost; and (ii) when there is a constraint on the average number of transmissions, what is the minimum estimation error. For both these cases, we characterize the optimal transmission and estimation strategies that achieve the optimal trade-off.

Two approaches have been used in the literature to investigate real-time or zero-delay communication. The first approach considers coding of individual sequences [1–4]; the second approach considers coding of Markov sources [5–10]. The model presented above fits with the latter approach. In particular, it may be viewed as real-time transmission, which is noiseless but expensive. In most of the results in the literature, the focus has been on identifying sufficient statistics (or information states) at the transmitter and the receiver; for some of the models, a dynamic programming decomposition has also been derived. However, very little is known about the solution of these dynamic programs.

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 state to the receiver. These transmitted events *reset* the system. We exploit these special features to identify an analytic solution to the dynamic program corresponding to the above communication system.

Several variations of the communication system described above have been considered in the literature. The most closely related models are [11–15] which are summarized below. Other related work includes censoring sensors [16,17] (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 [18–20] (where the receiver decides when the sensor should transmit), sensor sleep scheduling [21–24] (where the sensor is allowed to sleep for a pre-specified amount of time); and event-based communication [25–27] (where the sensor transmits when a certain event takes place). We contrast our model with [11–15] below.

In [11], the authors considered a remote estimation problem where the sensor could communicate a finite number of times. They assumed that the sensor used a threshold strategy to decide when to communicate and determined the optimal estimation strategy and the value of the thresholds.

In [12], the authors considered remote estimation of a Gauss-Markov process. They assumed a particular form of the estimator and showed that the estimation error is a sufficient statistic for the sensor.

In [13], the authors considered remote estimation of a scalar Gauss-Markov process but did not impose any assumption on the transmission or estimation strategy. They used ideas from majorization theory to show that the optimal estimation strategy is Kalman-like and the optimal transmission strategy is threshold based. The results of [13] were generalized to other setups in [14] and [15]. In [14], the authors considered remote estimation of countable state Markov processes where the sensor harvests energy to communicate. Similar to the approach taken in [13], the authors used majorization theory to show that if the Markov process is

driven by symmetric and unimodal noise process then the structural results of [13] continue to hold. In [15], the authors considered remote estimation of a scalar first-order autoregressive source. They used a person-by-person optimization approach to identify an iterative algorithm to compute the optimal transmission and estimation strategy. They showed that if the autoregressive process is driven by a symmetric unimodal noise process, then the iterative algorithm has a unique fixed point and the structural results of [13] continue to hold.

In all these papers [13–15], a dynamic program to compute the optimal thresholds was also identified. However, the problem of computing the optimal thresholds by solving the dynamic program was not investigated.

1.2 Contributions

We investigate remote estimation of two models of Markov processes—discrete symmetric Markov processes (Model A) and Gauss-Markov processes (Model B)—under two infinite horizon setups: the discounted setup with discount factor $\beta \in (0,1)$ and the long term average setup, which we denote by $\beta = 1$ for uniformity of notation. For both models, we consider two fundamental trade-offs:

1. *Costly communication*: When each transmissions costs λ units, what is the minimum achievable cost of communication plus estimation error, which we denote by $C^*_\beta(\lambda)$.
2. *Constrained communication*: When the average number of transmissions are constrained by α, what is the minimum achievable estimation error, which we denote by $D^*_\beta(\alpha)$ and refer to as the *distortion-transmission* trade-off.

We completely characterize both trade-offs. In particular,

- In Model A, $C^*_\beta(\lambda)$ is continuous, increasing, piecewise-linear, and concave in λ while $D^*_\beta(\alpha)$ is continuous, decreasing, piecewise-linear, and convex in α. We derive explicit expressions (in terms of simple matrix products) for the corner points of both these curves.
- In Model B, $C^*_\beta(\lambda)$ is continuous, increasing, and concave in λ while $D^*_\beta(\alpha)$ is continuous, decreasing, and convex in α. We characterize how these curve scale as a function of the noise variance σ^2 and show that they can be completely characterized by $C^*_\beta(\lambda)$ and $D^*_\beta(\alpha)$ for $\sigma = 1$. We derive an algorithmic procedure to compute the latter curves by using solutions of Fredholm integral equations of the second kind.

We also explicitly identify transmission and estimation strategies that achieve any point on these trade-off curves. For all cases, we show that we can restrict attention to time-homogeneous strategies in which the estimation decision is to choose the last transmitted symbol and the transmission decision is made by comparing the instantaneous estimation error when transmission is not made with a fixed threshold. In the constrained communication setup for Model A, the transmission strategy is a randomized strategy; in all other setups, the transmission strategy is a deterministic strategy. In addition,

- In Model A, the optimal threshold as a function of λ and α can be computed using a look-up table.
- In Model B, the optimal threshold as function of λ and α is an appropriately scaled version of the threshold for the case of $\sigma = 1$. For $\sigma = 1$, we derive an algorithmic procedure to compute the optimal threshold by using the solutions of Fredholm integral equations of the second kind.

1.3 Notation

Throughout this paper, we use the following notation. $\mathbb{Z}$, $\mathbb{Z}_{\geq 0}$ and $\mathbb{Z}_{>0}$ denote the set of integers, the set of non-negative integers and the set of strictly positive integers, respectively. Similarly, $\mathbb{R}$, $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ denote the set of reals, the set of non-negative reals and the set of strictly positive reals, 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

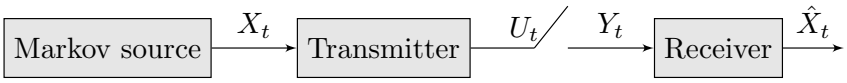


Figure 1: A block diagram depicting the communication system considered in this paper.

(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 $\{-k, \ldots, k\}$; 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.

2 Model and problem formulation

2.1 Model

Consider a discrete-time Markov process $\{X_t\}_{t=0}^{\infty}$ with initial state $X_0 = 0$ and for $t \geq 0$

$$X_{t+1} = X_t + W_t, \tag{1}$$

where $\{W_t\}_{t=0}^{\infty}$ is an i.i.d. noise process. We consider two specific models:

- **Model A:** $X_t, W_t \in \mathbb{Z}$ and W_t is distributed according to a unimodal and symmetric distribution p, i.e. for all $e \in \mathbb{Z}_{\geq 0}$, $p_e = p_{-e}$ and $p_e \geq p_{e+1}$. To avoid trivial cases, we assume $p_1 > 0$.
- **Model B:** $X_t, W_t \in \mathbb{R}$ and W_t is a zero-mean Gaussian random variable with variance σ^2. The pdf of W_t is denoted by $\phi(\cdot)$.

A sensor sequentially observes the process and at each time, chooses whether or not to transmit the current state. 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

$$U_t = f_t(X_{0:t}, U_{0:t-1}). \tag{2}$$

We use the short-hand notation $X_{0:t}$ to denote the sequence $(X_0, \ldots, X_t)$. 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 sequentially observes $\{Y_t\}_{t=0}^{\infty}$ and generates an estimate $\{\hat{X}_t\}_{t=0}^{\infty}$ (where $\hat{X}_t \in \mathbb{Z}$) using an *estimation strategy* $g = \{g_t\}_{t=0}^{\infty}$, i.e.,

$$\hat{X}_t = g_t(Y_{0:t}). \tag{3}$$

The fidelity of the estimation is measured by a per-step distortion $d(X_t - \hat{X}_t)$.

- For Model A, we assume that $d(0) = 0$, for $e \neq 0$, $d(e) \neq 0$ and that $d(\cdot)$ is even and increasing on $\mathbb{Z}_{\geq 0}$, i.e. for all $e \in \mathbb{Z}_{\geq 0}$, $d(e) = d(-e)$ and $d(e) \leq d(e + 1)$.
- For Model B, we assume that $d(e) = e^2$.

2.2 Performance measures

Given a transmission and estimation strategy (f, g) and a discount factor $\beta \in (0, 1]$, we define the expected distortion and the expected number of transmissions as follows. For $\beta \in (0, 1)$, the expected *discounted* distortion is given by

$$D_\beta(f, g) \coloneqq (1 - \beta)\mathbb{E}^{(f,g)}\Big[\sum_{t=0}^{\infty} \beta^t d(X_t - \hat{X}_t) \;\Big|\; X_0 = 0\Big] \tag{4}$$

and for $\beta = 1$, the expected *long-term average* distortion is given by

$$D_1(f, g) \coloneqq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{(f,g)}\Big[\sum_{t=0}^{T-1} d(X_t - \hat{X}_t) \;\Big|\; X_0 = 0\Big]. \tag{5}$$

Similarly, for $\beta \in (0, 1)$, the expected *discounted* number of transmissions is given by

$$N_\beta(f, g) \coloneqq (1 - \beta)\mathbb{E}^{(f,g)}\Big[\sum_{t=0}^{\infty} \beta^t U_t \;\Big|\; X_0 = 0\Big] \tag{6}$$

and for $\beta = 1$, the expected *long-term average* number of transmissions is given by

$$N_1(f, g) \coloneqq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{(f,g)}\Big[\sum_{t=0}^{T-1} U_t \;\Big|\; X_0 = 0\Big]. \tag{7}$$

2.3 Problem formulations

We are interested in the following two optimization problems.

Problem 1 (Costly communication) *In the model of Section 2.1, given a discount factor $\beta \in (0, 1]$ and a communication cost $\lambda \in \mathbb{R}_{>0}$, find a transmission and estimation strategy (f^*, g^*) such that*

$$C^*_\beta(\lambda) \coloneqq C_\beta(f^*, g^*; \lambda) = \inf_{(f,g)} C_\beta(f, g; \lambda), \tag{8}$$

where

$$C_\beta(f, g; \lambda) \coloneqq D_\beta(f, g) + \lambda N_\beta(f, g)$$

is the total communication cost and the infimum in (8) is taken over all history-dependent strategies.

Problem 2 (Constrained communication) *In the model of Section 2.1, given a discount factor $\beta \in (0, 1]$ and a costant $\alpha \in (0, 1)$, find a transmission and estimation strategy (f^*, g^*) such that*

$$D^*_\beta(\alpha) \coloneqq D_\beta(f^*, g^*) = \inf_{(f,g):N_\beta(f,g)\le\alpha} D_\beta(f, g), \tag{9}$$

where the infimum is taken over all history-dependent strategies.

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.*

2.4 Preliminary results

Proposition 2.1 *For any $\beta \in (0, 1]$ and $\lambda \ge 0$, $C^*_\beta(\lambda)$ is increasing and concave function of λ.*

Proof. Note that for any (f, g), the function $C_\beta(f, g; \lambda)$ is increasing (because $N_\beta(f, g) \ge 0$) and affine in λ. The infimum of increasing functions is increasing; hence, $C^*_\beta(\lambda)$ is increasing in λ. The infimum of affine functions is concave; hence $C^*_\beta(\lambda)$ is concave in λ. □

Proposition 2.2 *For any $\alpha \in (0,1)$, the distortion-transmission function $D^*(\alpha)$ is decreasing and convex function of α.*

Proof. $D^*_\beta(\alpha)$ is the solution to a constrained optimization problem and the constraint set $\{(f,g) : N_\beta(f,g) \leq \alpha\}$ increases with α. Hence, $D^*_\beta(\alpha)$ decreases with α. To see that $D^*_\beta(\alpha)$ is convex in α, consider $\alpha_1 < \alpha < \alpha_2$ and suppose (f_1, g_1) and (f_2, g_2) are optimal policies for α_1 and α_2 respectively. Let $\theta = (\alpha - \alpha_1)/(\alpha_2 - \alpha_1)$ and (f, g) be a mixed strategy that picks (f_1, g_1) with probability θ and (f_2, g_2) with probability $(1-\theta)$ (Note that the randomization is done only at the start of communication). Then $N_\beta(f,g) = \alpha$ and consequently, $D^*_\beta(\alpha) < D_\beta(f,g) = \theta D_\beta(f_1,g_1) + (1-\theta) D_\beta(f_2,g_2)$. Hence, $D^*_\beta(\alpha)$ is convex. □

Remark 2.1 It can be shown that $\lim_{\alpha\to 0} D^*_\beta(\alpha) = \infty$ [1] and $\lim_{\alpha\to 1} D^*_\beta(\alpha) = 0$.

2.5 Organization of the paper

In the rest of the paper, we completely characterize the functions $C^*_\beta(\lambda)$ and $D^*_\beta(\alpha)$. In Section 3 we discuss the structure of the optimal strategies; first for finite horizon setup, and then for infinite horizon setup. In Section 4 we provide some relevant definitions, properties and computations of some relevant parameters for both Models A and B, which lay the background to analyze the main results thereafter. We present the main results for Models A and B in Sections 5 and 6, respectively. Lastly, in Section 7 we validate the analytical results with an example for Model A and provide easily computable closed form expressions for all relevant parameters.

3 Structure of optimal strategies

3.1 Finite horizon setup

Finite horizon version of Problem 1 has been investigated in [14] (for Model A) and in [13,15] (for Model B), where the structure of the optimal transmission and estimation strategy was established. To describe these results, we define the following.

Definition 3.1 *Let Z_t denote the most recently transmitted value of the Markov source. The process $\{Z_t\}_{t=0}^\infty$ evolves in a controlled Markov manner as follows: $Z_0 = 0$, and*

$$Z_t = \begin{cases} X_t, & \text{if } U_t = 1; \\ Z_{t-1}, & \text{if } U_t = 0. \end{cases}$$

Note that since U_t can be inferred from the transmitted symbol Y_t, the receiver can also keep track of Z_t as follows: $Z_0 = 0$, and

$$Z_t = \begin{cases} Y_t, & \text{if } Y_t \neq \mathfrak{E}; \\ Z_{t-1}, & \text{if } Y_t = \mathfrak{E}. \end{cases}$$

Definition 3.2 *Let $E_t = X_t - Z_{t-1}$. The process $\{E_t\}_{t=0}^\infty$ evolves in the following manner*

$$E_{t+1} = (1 - U_t)E_t + W_t.$$

Note that the transmitter can keep track of E_t.

Remark 3.1 Note that each transmission *resets* the state of the error process to $w \in \mathbb{Z}$ with probability p_w. In between the transmissions, the error process evolves in a Markovian manner.

[1] A symmetric Markov chain defined over $\mathbb{Z}$ or $\mathbb{R}$ does not have a stationary distribution. Therefore, in the limit of no transmission, the expected distribution diverges to ∞.

Theorem 1 *For a finite horizon version of Problem 1, the processes $\{Z_t\}$ and $\{E_t\}$ are sufficient statistics at the estimator and the transmitter respectively. In particular, an optimal estimation strategy is given by*

$$\hat{X}_t = g_t^*(Z_t) = Z_t, \tag{10}$$

and an optimal transmission strategy is given by

$$U_t = f_t(E_t) = \begin{cases} 1, & \text{if } |E_t| \geq k_t; \\ 0, & \text{if } |E_t| < k_t. \end{cases} \tag{11}$$

The above structural results were obtained in [14, Theorems 2 and 3] for Model A and in [13, Theorm 1] and [15, Lemmas 1, 3 and 4] for Model B.

Remark 3.2 The results in [14] were derived under the assumption that $\{W_t\}$ has finite support. These results can be generalized for $\{W_t\}$ having countable support using ideas from [28]. For that reason, we state Theorem 1 without any restriction on the support of $\{W_t\}$. See the supplementary document for the generalization of [14, Theorems 2 and 3] to $\{W_t\}$ with countable support.

Remark 3.3 In general, the optimal estimation strategy depends on the choice of the transmission strategy and vice-versa. Theorem 1 shows that when the noise process and the distortion function satisfy appropriate symmetry assumptions, the optimal estimation strategy can be specified in closed form. Consequently, we can fix the receiver to be of the above form, and consider the centralized problem of identifying the best transmission strategy.

3.2 Infinite horizon setup and the structure of optimal strategies

As explained in Remark 3.3, we can fix the estimation strategy and find the transmission strategy that is the *best response* to this estimation strategy. Identifying such a best response strategy is a centralized stochastic control problem. Since the optimal estimation strategy is time-homogeneous, one expects the optimal transmission strategy (i.e., the choice of the optimal thresholds $\{k_t\}_{t=0}^{\infty}$) to be time-homogeneous as well. To establish such a result, we need the following technical assumption for Model A.

(A1) For every $\lambda \geq 0$, there exists a function $\rho : \mathbb{Z} \to \mathbb{R}$ and positive and finite constants μ_1 and μ_2 such that for all $e \in \mathbb{Z}$, we have that $\max\{\lambda, d(e)\} \leq \mu_1 \rho(e)$, and

$$\max\left\{\sum_{n=-\infty}^{\infty} p_{n-e}\rho(n), \sum_{n=-\infty}^{\infty} p_n \rho(n)\right\} \leq \mu_2 \rho(e).$$

Theorem 2 *Consider Problem 1 for $\beta \in (0,1]$ and an estimation strategy given by (10). Assume that Assumption (A1) is satisfied for Model A. Then, an optimal transmission strategy (for both Models A and B) is of the form*

$$U_t = f(E_t) = \begin{cases} 1, & \text{if } |E_t| \geq k; \\ 0, & \text{if } |E_t| < k, \end{cases} \tag{12}$$

where the threshold k is time-homogeneous.

3.3 Proof of Theorem 2

To prove the result, we proceed as follows:

1. We show that the result of the theorem is true for $\beta \in (0,1)$ and the optimal strategy is given by an appropriate dynamic program.
2. We show that the value function of the dynamic program is even and increasing on $\mathbb{Z}_{\geq 0}$ (for Model A) and even and increasing on $\mathbb{R}_{\geq 0}$ (for Model B).

3. For $\beta = 1$, we use the vanishing discount approach to show that the optimal strategy for the long-term average cost setup may be determined as a limit to the optimal strategy for the discounted cost setup is the discount factor $\beta \uparrow 1$.

3.3.1 The discounted setup

(a) Model A: The optimal transmission strategy is given by the solution (if it exists) of the following dynamic program: for all $e \in \mathbb{Z}$,

$$V_\beta(e;\lambda) = \min_{u\in\{0,1\}} \Big[c(e,u) + \beta \sum_{w\in\mathbb{Z}} p_w V_\beta(e+w;\lambda)\Big], \tag{13}$$

where $c(e,u) = \lambda u + (1-u)d(e)$ is the one-stage cost. It follows from [29, Proposition 6.10.3] that the above dynamic program has a unique bounded solution. Note that Assumption (A1) is equivalent to [29, Assumptions 6.10.1, 6.10.2] used in [29, Proposition 6.10.3].

(b) Model B: As for Model A, the optimal transmission strategy is given by the solution (if it exists) of the following dynamic program: for all $e \in \mathbb{R}$,

$$V_\beta(e;\lambda) = \min_{u\in\{0,1\}} \Big[c(e,u) + \beta \int_{\mathbb{R}} \phi(w) V_\beta(e+w;\lambda) dw\Big], \tag{14}$$

where $c(e,u) = \lambda u + (1-u)d(e)$ is the one-stage cost. It follows from [30, Theorem 4.2.3] that the above dynamic program has a unique bounded solution because: (i) $c(\cdot,\cdot)$ and $\phi(\cdot)$ satisfy [30, Assumption 4.2.1] and (ii) the cost of the 'always transmit' policy is λ, hence [30, Assumption 4.2.2] is satisfied.

3.3.2 Properties of the value function

Proposition 3.1 *For any $\lambda > 0$, the value functions $V_\beta(\cdot;\lambda)$ given by (13) and (14) are even and increasing on $\mathbb{Z}_{\geq 0}$ and on $\mathbb{R}_{\geq 0}$, respectively.*

See Appendix A for the proof of Proposition 3.1.

3.3.3 The long-term average setup

Proposition 3.2 *For any $\lambda \geq 0$, the value function $V_\beta(\cdot;\lambda)$ for Models A and B, as given by (13) and (14) respectively, satisfy the following SEN conditions of [30, 31]:*

(S1) There exists a reference state $e_0 \in \mathbb{Z}$ for Model A and $e_0 \in \mathbb{R}$ for Model B and a non-negative scalar M_λ such that $V_\beta(e_0,\lambda) < M_\lambda$ for all $\beta \in (0,1)$.

(S2) Define $h_\beta(e;\lambda) = (1-\beta)^{-1}[V_\beta(e;\lambda) - V_\beta(e_0;\lambda)]$. There exists a function $K_\lambda : \mathbb{Z} \to \mathbb{R}$ such that $h_\beta(e;\lambda) \leq K_\lambda(e)$ for all $e \in \mathbb{Z}$ for Model A and for all $e \in \mathbb{R}$ for Model B and $\beta \in (0,1)$.

(S3) There exists a non-negative (finite) constant L_λ such that $-L_\lambda \leq h_\beta(e;\lambda)$ for all $e \in \mathbb{Z}$ for Model A and for all $e \in \mathbb{R}$ for Model B and $\beta \in (0,1)$.

Therefore, if f_β denotes an optimal strategy for $\beta \in (0,1)$, and f_1 is any limit point of $\{f_\beta\}$, then f_1 is optimal for $\beta = 1$.

Proof. Let $V_\beta^{(0)}(e,\lambda)$ denote the value function of the 'always transmit' strategy. Since $V_\beta(e,\lambda) \leq V_\beta^{(0)}(e,\lambda)$ and $V_\beta^{(0)}(e,\lambda) = \lambda$, (S1) is satisfied with $M_\lambda = \lambda$.

We show (S2) for Model B, but a similar argument works for Model A as well. Since not transmitting is optimal at state 0, we have

$$V_\beta(0,\lambda) = \beta \int_{-\infty}^{\infty} \phi(w) V_\beta(w,\lambda) dw.$$

Let $V_\beta^{(1)}(e,\lambda)$ denote the value function of the strategy that transmits at time 0 and follows the optimal strategy from then on. Then

$$\begin{aligned} V_\beta^{(1)}(e,\lambda) &= (1-\beta)\lambda + \beta \int_{-\infty}^{\infty} \phi(w) V_\beta(w,\lambda) dw \\ &= (1-\beta)\lambda + \beta V_\beta(0,\lambda) \end{aligned} \tag{15}$$

Since $V_\beta(e,\lambda) \leq V_\beta^{(1)}(e,\lambda)$, from (15) we get that $(1-\beta)^{-1}[V_\beta(e,\lambda) - V_\beta(0,\lambda)] \leq \lambda$. Hence (S2) is satisfied with $K_\lambda(e) = \lambda$.

By Proposition 3.1, $V_\beta(e,\lambda) \geq V_\beta(0,\lambda)$, hence (S3) is satisfied with $L_\lambda = 0$. □

Proof of Theorem 2. Since the value function $V_\beta(\cdot,\lambda)$ satisfies the SEN conditions for reference state $e_0 = 0$, the optimaity of the threshold strategy for long-term average setup follows from [31, Theorem 7.2.3] for Model A and [30, Theorem 5.4.3] for Model B, respectively. □

We are now ready to provide the analytical results for both Models A and B. Before we go into the detailed discussions about the main results, we lay a background in the form of definitions, properties and computations of some relevant parameters in the next section to facilitate easy comprehension.

4 Preliminary results

4.1 Some definitions

Let $\mathcal{F}$ denote the class of all time-homogeneous threshold-based strategies of the form (12). Let $f^{(k)} \in \mathcal{F}$ denote the strategy with threshold k, $k \in \mathbb{Z}_{\geq 0}$, i.e.,

$$f^{(k)}(e) := \begin{cases} 1, & \text{if } |e| \geq k; \\ 0, & \text{if } |e| < k. \end{cases}$$

When the system starts in state e and follows strategy $f^{(k)}$, define for $\beta \in (0,1]$ the following (where $\beta \in (0,1)$ implies the discounted cost setup and $\beta = 1$ implies long-term average cost setup):

- $L_\beta^{(k)}(e)$: the expected distortion until the first transmission.
- $M_\beta^{(k)}(e)$: the expected time until the first transmission
- $D_\beta^{(k)}(e)$: the expected distortion
- $N_\beta^{(k)}(e)$: the expected number of transmissions
- $C_\beta^{(k)}(e;\lambda)$: the expected total cost, i.e.,

$$C_\beta^{(k)}(e;\lambda) = D_\beta^{(k)}(e) + \lambda N_\beta^{(k)}(e), \quad \lambda \geq 0.$$

Note that $D_\beta^{(k)}(0) = D_\beta(f^{(k)}, g^*)$, $N_\beta^{(k)}(0) = N_\beta(f^{(k)}, g^*)$ and $C_\beta^{(k)}(0;\lambda) = C_\beta(f^{(k)}, g^*; \lambda)$.

Let the stopping set $S^{(k)}$ be defined as the following

$$S^{(k)} := \begin{cases} \{-(k-1), \cdots, k-1\}, & \text{for Model A;} \\ (-k, k), & \text{for Model B.} \end{cases}$$

Define operators $\mathcal{B}$ and $\mathcal{B}^{(k)}$ as follows:

- **Model A**: For any $v : \mathbb{Z} \to \mathbb{R}$, define operator $\mathcal{B}$ as

$$[\mathcal{B}v](e) \coloneqq \sum_{w=-\infty}^{\infty} p_w v(e+w), \quad \forall e \in \mathbb{Z}.$$

  Note that an equivalent definition is

$$[\mathcal{B}v](e) \coloneqq \sum_{n=-\infty}^{\infty} p_{n-e} v(n), \quad \forall e \in \mathbb{Z}.$$

  Furthermore, for any $v^{(k)} : S^{(k)} \to \mathbb{R}$, define operator $\mathcal{B}^{(k)}$ as

$$[\mathcal{B}^{(k)}v](e) \coloneqq \sum_{n \in S^{(k)}} p_{n-e} v(n), \quad \forall e \in S^{(k)}.$$

- **Model B**: For any bounded $v : \mathbb{R} \to \mathbb{R}$, define operator $\mathcal{B}$ as

$$[\mathcal{B}v](e) \coloneqq \int_{\mathbb{R}} \phi(w) v(e+w) dw, \quad \forall e \in \mathbb{R}.$$

  Or, equivalently,

$$[\mathcal{B}v](e) \coloneqq \int_{\mathbb{R}} \phi(n-e) v(n) dn, \quad \forall e \in \mathbb{R}.$$

  Furthermore, for any $v^{(k)} : S^{(k)} \to \mathbb{R}$, define operator $\mathcal{B}^{(k)}$ as

$$[\mathcal{B}^{(k)}v](e) \coloneqq \int_{S^{(k)}} \phi(n-e) v(n) dn, \quad \forall e \in S^{(k)}.$$

Let $\|\cdot\|_\infty$ denote the sup-norm, i.e. for any $v : S^{(k)} \to \mathbb{R}$,

$$\|v\|_\infty = \sup_{e \in S^{(k)}} |v(e)|.$$

Lemma 4.1 *In both Model A and B, the operator $\mathcal{B}^{(k)}$ is contraction, i.e., for any $v : S^{(k)} \to \mathbb{R}$,*

$$\|\mathcal{B}^{(k)}v\|_\infty < \|v\|_\infty.$$

Thus, for any bounded $h : S^{(k)} \to \mathbb{R}$, the equation

$$v = h + \beta \mathcal{B}^{(k)} v \tag{16}$$

has a unique bounded solution v. In addition, if h is continuous, then v is continuous.

Proof. We state the proof for Model B. The proof for Model A is similar. By the definition of sup-norm, we have that for any bounded h

$$\begin{aligned}
\|\mathcal{B}^{(k)}v\|_\infty &= \sup_{e \in [-k,k]} \int_{-k}^{k} \phi(w-e) v(w) dw \\
&\leq \sup_{e \in [-k,k]} \beta \|v\|_\infty \int_{-k}^{k} \phi(w-x) dw \\
&< \|v\|_\infty, \quad \text{(since } \phi \text{ is a pdf).}
\end{aligned}$$

Now, consider the operator $\mathcal{B}'$ given as: $\mathcal{B}'v = h + \mathcal{B}^{(k)}v$. Then we have,

$$\|\mathcal{B}'(v_1 - v_2)\|_\infty = \|\mathcal{B}^{(k)}(v_1 - v_2)\|_\infty < \|v_1 - v_2\|_\infty.$$

Since the space of bounded real-valued functions is compact, by Banach fixed point theorem, $\mathcal{B}'$ has a unique fixed point.

If h is continuous, we can define $\mathcal{B}^{(k)}$ and $\mathcal{B}'$ as operators on the space of continuous and bounded real-valued function (which is compact). Hence, the continuity of the fixed point follows also from Banach fixed point theorem. □

4.2 Expressions for $L_\beta^{(k)}$ and $M_\beta^{(k)}$

Recall from Remark 3.1 that the state E_t evolves in a Markovian manner until the first transmission. We may equivalently consider the Markov process until it is absorbed in $(-\infty, -k] \cup [k, \infty)$. Thus, from balance equation for Markov processes, we have for all $e \in S^{(k)}$,

$$L_\beta^{(k)}(e) = d(e) + \beta[\mathcal{B}^{(k)} L_\beta^{(k)}](e), \tag{17}$$

$$M_\beta^{(k)}(e) = 1 + \beta[\mathcal{B}^{(k)} M_\beta^{(k)}](e). \tag{18}$$

Lemma 4.2 *For any $\beta \in (0, 1]$, equations (17) and (18) have unique and bounded solutions that are*

(a) strictly increasing in k,

(b) continuous and differentiable in k for Model B,

(c) $\lim_{\beta \uparrow 1} L_\beta^{(k)}(e) = L_1^{(k)}(e)$, for all e.

Proof. The solutions of equations (17) and (18) exist due to Lemma 4.1.

(a) Consider k, l (in $\mathbb{Z}_{\geq 0}$ for Model A and in $\mathbb{R}_{\geq 0}$ for Model B) such that $k < l$. A sample path starting from $e \in S^{(k)}$ must escape $S^{(k)}$ before it escapes $S^{(l)}$. Thus

$$L_\beta^{(l)}(e) \geq L_\beta^{(k)}(e).$$

In addition, the above inequality is strict because W_t has a unimodal distribution.

(b) The continuity and differentiability can be proved from elementary algebra. See Appendix B in the supplementary document for details.

(c) The limit holds since $L_\beta^{(k)}(e)$ and $M_\beta^{(k)}(e)$ are continuous functions of β.

□

4.3 Computation of $L_\beta^{(k)}$ and $M_\beta^{(k)}$

For Model A, the values of $L_\beta^{(k)}$ and $M_\beta^{(k)}$ can be computed by observing that the operator $\mathcal{B}^{(k)}$ is equivalent to a matrix multiplication. In particular, define the matrix $P^{(k)}$ as

$$P_{ij}^{(k)} := p_{|i-j|}, \quad \forall i, j \in S^{(k)}.$$

Then,

$$\begin{aligned}[\mathcal{B}^{(k)} v](e) &= \sum_{n \in S^{(k)}} p_{n-e} v(n) \\ &= \sum_{n \in S^{(k)}} P_{n-e}^{(k)} v(n) \\ &= [P^{(k)} v]_e \end{aligned} \tag{19}$$

With a slight abuse of notation, we are using v both as a function and a vector. For ease of notation, define the matrix $Q^{(k)}$ and the vector $d^{(k)}$ as follows:

$$Q_\beta^{(k)} := [I_{2k-1} - \beta P^{(k)}]^{-1}, \tag{20}$$

$$d^{(k)} := [d(-k+1), \ldots, d(k-1)]^\intercal. \tag{21}$$

Then, an immediate consequence of (19), (17) and (18) is the following:

Proposition 4.1 *In Model A, for any $\beta \in (0,1]$,*

$$L_\beta^{(k)} = [I_{2k-1} - \beta P^{(k)}]^{-1} d^{(k)} \tag{22}$$

$$M_\beta^{(k)} = [I_{2k-1} - \beta P^{(k)}]^{-1} \mathbf{1}_{2k-1}. \tag{23}$$

For Model B, for any $\beta \in (0,1]$, (17) and (18) are Fredholm integral equations of second kind [32]. The solution can be computed by identifying the inverse operator

$$\mathcal{Q}_\beta^{(k)} = [I - \beta \mathcal{B}^{(k)}]^{-1}, \tag{24}$$

which is given by

$$[\mathcal{Q}_\beta^{(k)} v](e) = \int_{-k}^{k} R_\beta^{(k)}(e,w) v(w) dw, \tag{25}$$

where $R_\beta^{(k)}(\cdot,\cdot)$ is the resolvent of ϕ and can be computed using the Liouville-Neumann series. See [32] for details. Since ϕ is smooth, (17) and (18) can also be solved by discretizing the integral equation. A Matlab implementation of this approach is presented in [33].

4.4 Expressions for $D_\beta^{(k)}$ and $N_\beta^{(k)}$, $\beta \in (0,1]$

As discussed in Remark 3.1, the error process $\{E_t\}_{t=0}^\infty$ is a controlled Markov process. Therefore, the functions $D_\beta^{(k)}$ and $N_\beta^{(k)}$ may be thought as value functions when strategy $f^{(k)}$ is used. Thus, they satisfy the following fixed point equations: for $\beta \in (0,1)$,

$$D_\beta^{(k)}(e) = \begin{cases} \beta[\mathcal{B} D_\beta^{(k)}](0), & \text{if } |e| \geq k \\ (1-\beta) d(e) + \beta[\mathcal{B} D_\beta^{(k)}](e), & \text{if } |e| < k, \end{cases} \tag{26}$$

$$N_\beta^{(k)}(e) = \begin{cases} (1-\beta) + \beta[\mathcal{B} N_\beta^{(k)}](0), & \text{if } |e| \geq k \\ \beta[\mathcal{B} N_\beta^{(k)}](e), & \text{if } |e| < k. \end{cases} \tag{27}$$

Proposition 4.2 *There exists unique and bounded functions $D_\beta^{(k)}(e)$ and $N_\beta^{(k)}(e)$ that satisfy (26) and (27), are even and increasing (on $\mathbb{Z}_{\geq 0}$ for Model A and on $\mathbb{R}_{\geq 0}$ for Model B) in e, and satisfy the SEN conditions. Thus,*

$$D_1^{(k)}(e) = \lim_{\beta \uparrow 1} D_\beta^{(k)}(e) \quad \text{and} \quad N_1^{(k)}(e) = \lim_{\beta \uparrow 1} N_\beta^{(k)}(e).$$

The proof follows from the arguments similar to those of Section 3.3 and is omitted.

Using (26) and (27), the performance of strategy $(f^{(k)}, g^*)$ is given as follows:

Proposition 4.3 *For any $\beta \in (0,1]$, the performance of strategy $f^{(k)}$ for the discounted case of costly communication in both Models A and B is given as follows:*

1. *$D_\beta(f^{(0)}, g^*) = 0$, $N_\beta(f^{(0)}, g^*) = 1$, and $C_\beta(f^{(0)}, g^*; \lambda) = \lambda$.*
2. *For $k \in \mathbb{Z}_{>0}$*

$$D_\beta(f^{(k)}, g^*) = \frac{L_\beta^{(k)}(0)}{M_\beta^{(k)}(0)},$$

$$N_\beta(f^{(k)}, g^*) = \frac{1}{M_\beta^{(k)}(0)} - (1-\beta),$$

and

$$C_\beta(f^{(k)}, g^*; \lambda) = \frac{L_\beta^{(k)}(0) + \lambda}{M_\beta^{(k)}(0)} - \lambda(1-\beta).$$

The proof is given in Appendix B.

Corollary 4.1 *In Model A, for any $\beta \in (0,1]$,*

$$D_\beta(f^{(1)}, g^*) = 0, \quad \text{and} \quad N_\beta(f^{(1)}, g^*) = \beta(1-p_0).$$

Lemma 4.3 *For both Models A and B, $D_\beta^{(k)}(e)$ is increasing in k for all e and all $\beta \in (0,1]$. When $\beta \in (0,1)$, the monotonicity is strict.*

See Appendix C for the proof.

4.5 Some additional properties of $C_\beta^*(\lambda)$

Lemma 4.4 *For both Models A and B, $C_\beta^{(k)}(\lambda)$ is submodular in (k,λ), i.e., for $l > k$, $C_\beta^{(l)}(0;\lambda) - C_\beta^{(k)}(0;\lambda)$ is decreasing in λ.*

Proof. $C_\beta^{(l)}(0;\lambda) - C_\beta^{(k)}(0;\lambda) = (D_\beta^{(l)}(0) - D_\beta^{(k)}(0)) - \lambda(N_\beta^{(k)}(0) - N_\beta^{(l)}(0))$. By Lemma 4.2 and Proposition 4.3, $N_\beta^{(k)}(0) - N_\beta^{(l)}(0)$ is positive, hence $C_\beta^{(l)}(0;\lambda) - C_\beta^{(k)}(0;\lambda)$ is increasing in λ. Hence $C_\beta^{(k)}(\lambda)$ is submodular. □

Proposition 4.4 *For both Models A and B,*

1. *Let $k_\beta^*(\lambda) = \arg\inf_{k\geq 0} C_\beta^{(k)}(0;\lambda)$ be the optimal k for a fixed λ. Then $k_\beta^*(\lambda)$ is increasing in λ.*
2. *In addition to being concave and increasing, $C_\beta^*(\lambda)$ is continuous in λ.*

Proof. We prove the result for Model B. Almost the same argument applies to Model A. Note that instead of optimizing over $k \geq 0$, we can restrict k to a compact set $[0, k^\circ]$, where $k^\circ := \min\{k\colon D_\beta^{(k)} > \lambda\}$. Such a k° always exists because $D_\beta^{(k)}$ is increasing in k (Lemma IV.3), $D_\beta^{(0)} = 0$ and $D_\beta^{(\infty)} = \infty$.

Any $k > k^\circ$ cannot be optimal because, $C_\beta^{(k)}(0;\lambda) > \lambda = C_\beta^{(0)}(0;\lambda)$. Hence, we can restrict k to the compact set $[0,k]$.

1. Since $k_\beta^*(\lambda)$ is the argmin of a submodular function over a compact set, it is increasing [34, Theorem 2.8.2].
2. Since $C_\beta^*(\lambda)$ is the pointwise minimum of a continuous function over a compact set, it is continuous.

□

5 Main results for Model A

5.1 Result for Problem 1

For Model A, $k_\beta^*(\lambda)$ takes values in $\mathbb{Z}_{\geq 0}$. Proposition 4.4 implies that $k_\beta^*(\lambda)$ is increasing; hence it must have a staircase shape as shown in Fig. 2. Let $\Lambda_\beta^{(k)}$ be the set of λ for which the strategy $f_\beta^{(k)}$ is optimal, i.e., for any $k \in \mathbb{Z}_{\geq 0}$,

$$\Lambda_\beta^{(k)} = \{\lambda \in \mathbb{R}_{\geq 0} : k_\beta^*(\lambda) = k\}. \tag{28}$$

Since $k_\beta^*(\lambda)$ has a staircase structure shown in Fig. 2, $\Lambda_\beta^{(k)}$ is an interval which we denote by $[\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$ and $\{\lambda_\beta^{(k)}\}_{k=0}^\infty$ is an increasing sequence. By continuity of $C_\beta^*(\lambda)$, we have that

$$C_\beta^{(k)}(0;\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(0;\lambda_\beta^{(k)}). \tag{29}$$

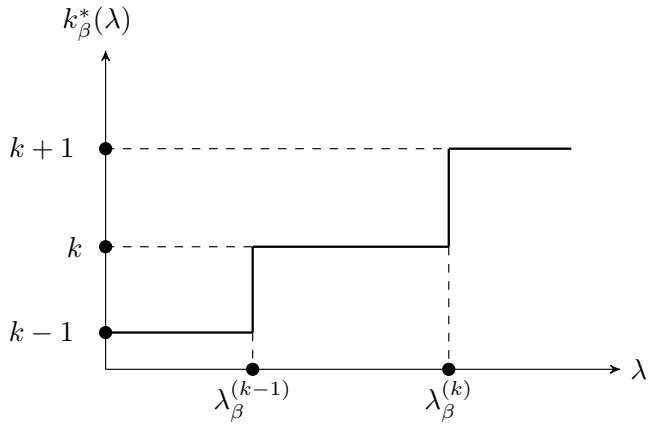


Figure 2: Plot of $k^*_\beta(\lambda)$ for Model A.

Substituting in the expression for $C_\beta^{(k)}$ from Proposition 4.3, we get that

$$\lambda_\beta^{(k)} = \frac{D_\beta^{(k+1)}(0) - D_\beta^{(k)}(0)}{N_\beta^{(k)}(0) - N_\beta^{(k+1)}(0)}. \tag{30}$$

By Lemma 4.2, both the numerator and the denominator are positive and, hence, $\lambda_\beta^{(k)}$ exists and is positive; see Fig. 3a–3b for illustration. Combining the above, the optimal strategy can be characterized as follows.

Theorem 3 *For all $\beta \in (0,1]$, we have the following.*

1. *For any $k \in \mathbb{Z}_{\geq 0}$ and any $\lambda \in [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$, the strategy $f^{(k)}$ is optimal for Problem 1.*
2. *The optimal performance for costly communication, $C^*_\beta(\lambda)$, in addition to being continuous, concave and increasing function of λ, is piecewise linear in λ.*

Proof. The proof of part 1) is an immediate consequence of the definition of (28) and 30. The piecewise linearity of C^*_β follows from part 1) and Proposition 4.3. □

5.2 Result for Problem 2

To describe the solution of Problem 2, we first define Bernoulli randomized strategy and Bernoulli randomized simple strategy.

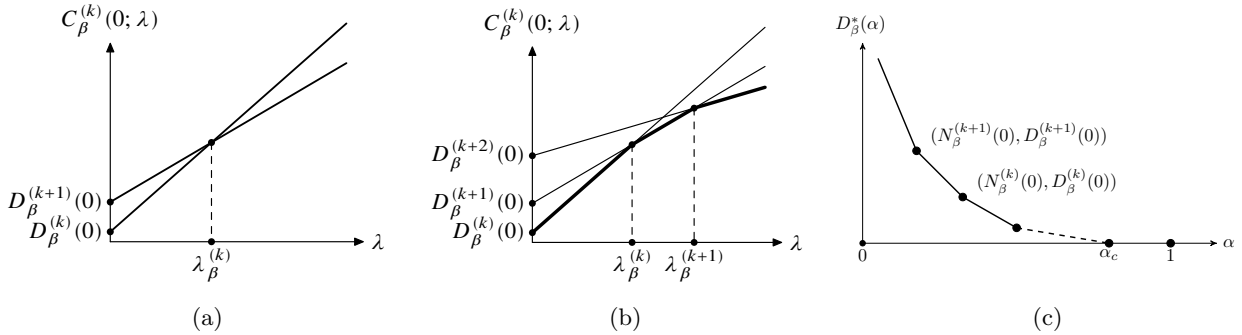


Figure 3: In Model A, (a) $\lambda_\beta^{(k)}$ is the abscissa of the intersection of $C_\beta^{(k)}(0;\lambda)$ and $C_\beta^{(k+1)}(0;\lambda)$; (b) $C^*_\beta(\lambda)$, which is shown in bold, is the minimum of $\{C_\beta^{(k)}(0;\lambda)\}_{k=0}^{\infty}$; (c) The distortion-transmission function $D^*_\beta(\alpha)$.

Definition 5.1 *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

$$k^*_\beta(\alpha) = \sup\{k \in \mathbb{Z}_{\geq 0} : N_\beta(f^{(k)}, g^*) \geq \alpha\} \tag{31}$$

and

$$\theta^*_\beta(\alpha) = \frac{\alpha - N_\beta(f^{(k^*_\beta(\alpha)+1)}, g^*)}{N_\beta(f^{(k^*_\beta(\alpha))}, g^*) - N_\beta(f^{(k^*_\beta(\alpha)+1)}, g^*)}. \tag{32}$$

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

$$\theta^* N_\beta(f^{(k^*)}, g^*) + (1-\theta^*) N_\beta(f^{(k^*+1)}, g^*) = \alpha. \tag{33}$$

Note that k^* and θ^* could have been equivalently defined as follow:

$$k^* = \sup\left\{k \in \mathbb{Z}_{\geq 0} : M^{(k)}_\beta \leq \frac{1}{1+\alpha-\beta}\right\},$$
$$\theta^* = \frac{M^{(k^*+1)} - \frac{1}{1+\alpha-\beta}}{M^{(k^*+1)} - M^{(k^*)}}.$$

Theorem 4 *Let f^* be the Bernoulli randomized simple strategy $(f^{(k^*)}, f^{(k^*+1)}, \theta^*)$, i.e.,*

$$f^*(e) = \begin{cases} 0, & \text{if } |e| < k^*; \\ 0, & \text{w.p. } 1-\theta^*, \text{ if } |e| = k^*; \\ 1, & \text{w.p. } \theta^*, \text{ if } |e| = k^*; \\ 1, & \text{if } |e| > k^*. \end{cases} \tag{34}$$

Then (f^*, g^*) is optimal for the constrained Problem (2) when $\beta \in (0,1]$.

Proof. The proof relies on the following characterization of the optimal strategy stated in [35, 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 [36]. Also, see [37, Theorem 8.4.1] 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 $\lambda^\circ \geq 0$ 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 $\lambda^\circ = \lambda^{(k^*)}_\beta$.

(f^*, g^*) satisfy (C1) due to (33). For $\lambda = \lambda^{(k^*)}_\beta$, 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 [35, Proposition 1.2], (f^*, g^*) is optimal for Problem 2. □

Theorem 5 *The distortion-transmission function is given by*

$$D^*_\beta(\alpha) = \theta^* D_\beta(f^{(k^*)}, g^*) + (1-\theta^*) D_\beta(f^{(k^*+1)}, g^*). \tag{35}$$

Furthermore, The distortion-transmission function, $D^*_\beta(\alpha)$, in addition to being continuous, convex and decreasing function of α, is piecewise linear in α.

Proof. The form of $D^*_\beta(\alpha)$ given in (35) follows immediately from the fact that (f^*, g^*) is a Bernoulli randomized simple strategy. To prove piecewise linearity of $D^*_\beta(\alpha)$, for every $k \in \mathbb{Z}_{\geq 0}$, define

$$\alpha^{(k)} = N_\beta(f^{(k)}, g^*),$$

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 (32)

$$\theta^* = \frac{\alpha - \alpha^{(k+1)}}{\alpha^{(k)} - \alpha^{(k+1)}},$$

and by (35),

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

Recall that $\alpha \in (\alpha^{(k+1)}, \alpha^{(k)})$ and, therefore, $D^*_\beta(\alpha)$ is piecewise linear. □

It follows from the argument given in the proof above that $\{(\alpha^{(k)}, D^{(k)}_\beta(0)\}_{k=0}^\infty$ are the vertices of the piecewise linear function D^*_β. See Fig. 3c for an illustration.

Combining Theorem 4 with the result of Corollary 4.1, we get

Corollary 5.1 *For any $\beta \in (0, 1]$,*

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

5.3 Discussion on deterministic implementation

The optimal strategy shown in Theorem 4 chooses a randomized action in states $\{-k^*, k^*\}$. It is also possible to identify deterministic (non-randomized) but time-varying strategies that achieve the same performance. We describe two such strategies for the long-term average setup.

5.3.1 Steering strategies

Let a^0_t (respectively, a^1_t) denote the number of times the action $u_t = 0$ (respectively, the action $u_t = 1$) has been chosen in states $\{-k^*, k^*\}$ in the past, i.e.

$$a^i_t = \sum_{s=0}^{t-1} \mathbb{1}\{|E_s| = k^*, u_s = i\}, \quad i \in \{0, 1\}.$$

Thus, the empirical frequency of choosing action $u_t = i$, $i \in \{0, 1\}$, in states $\{-k^*, k^*\}$ is $a^i_t/(a^0_t + a^1_t)$. A steering strategy compares these empirical frequencies with the desired randomization probabilities $\theta^0 = 1-\theta^*$

and $\theta^1 = \theta^*$ and chooses an action that *steers* the empirical frequency closer to the desired randomization probability. More formally, at states $\{-k^*, k^*\}$, the steering transmission strategy chooses the action

$$\arg\max_i \left\{\theta^i - \frac{a_t^i + 1}{a_t^0 + a_t^1 + 1}\right\}$$

in states $\{-k^*, k^*\}$ and chooses deterministic actions according to f^* (given in (34)) in states except $\{-k^*, k^*\}$. Note that the above strategy is deterministic (non-randomized) but depends on the history of visits to states $\{-k^*, k^*\}$. Such strategies were proposed in [38], where it was shown that the steering strategy descibed above achieves the same performance as the randomized startegy f^* and hence is optimal for Problem 2 for $\beta = 1$. Variations of such steering strategies have been proposed in [39, 40], where the adaptation was done by comparing the sample path average cost with the expected value (rather than by comparing empirical frequencies).

5.3.2 Time-sharing strategies

Define a cycle to be the period of time between consecutive visits of process $\{E_t\}_{t=0}^{\infty}$ to state zero. A time-sharing strategy is defined by a series $\{(a_m, b_m)\}_{m=0}^{\infty}$ and uses startegy $f^{(k^*)}$ for the first a_0 cycles, uses startegy $f^{(k^*+1)}$ for the next b_0 cycles, and continues to alternate between using startegy $f^{(k^*)}$ for a_m cycles and strategy $f^{(k^*+1)}$ for b_m cycles. In particular, if $(a_m, b_m) = (a, b)$ for all m, then the time-sharing strategy is a periodic strategy that uses $f^{(k^*)}$ a cycles and $f^{(k^*+1)}$ for b cycles.

The performance of such time-sharing strategies was evaluated in [41], where it was shown that if the cycle-lengths of the time-sharing strategy are chosen such that,

$$\begin{aligned}\lim_{M\to\infty} \frac{\sum_{m=0}^{M} a_m}{\sum_{m=0}^{M}(a_m + b_m)} &= \frac{\theta^* N_1^{(k^*)}}{\theta^* N_1^{(k^*)} + (1-\theta^*) N_1^{(k^*+1)}} \\ &= \frac{\theta^* N_1^{(k^*)}}{\alpha}.\end{aligned}$$

Then the time-sharing strategy $\{(a_m, b_m)\}_{m=0}^{\infty}$ achieves the same performance as the randomized strategy f^* and hence, is optimal for Problem 2 for $\beta = 1$.

6 Main results for Model B

6.1 Result for Problem 1

An immediate consequence of Lemma 4.2 and Proposition 4.3 is the following:

Corollary 6.1 *For any $\beta \in (0, 1]$, $D_\beta^{(k)}(0)$ and $N_\beta^{(k)}(0)$ are continuous in k. Furthermore, $N_\beta^{(k)}(0)$ is strictly decreasing in k.*

The differentiability of $L_\beta^{(k)}$ and $M_\beta^{(k)}$ in k (Lemma 4.2) and the expressions for $D_\beta^{(k)}$, $N_\beta^{(k)}$ and $C_\beta^{(k)}$ in Proposition 4.3 imply the following:

Corollary 6.2 *$D_\beta^{(k)}$, $N_\beta^{(k)}$ and $C_\beta^{(k)}$ are differentiable in k.*

We use $\partial_k D_\beta^{(k)}$, $\partial_k N_\beta^{(k)}$ and $\partial_k C_\beta^{(k)}$ to denote the derivative of $D_\beta^{(k)}$, $N_\beta^{(k)}$ and $C_\beta^{(k)}$ with respect to k.

Theorem 6 *For $\beta \in (0, 1]$, if the pair (λ, k) satisfy the following*

$$\lambda = -\frac{\partial_k D_\beta^{(k)}(0)}{\partial_k N_\beta^{(k)}(0)}, \tag{36}$$

then the strategy $(f^{(k)}, g^*)$ is λ-optimal for Problem 1. Furthermore, for any $k > 0$, there exists a $\lambda \geq 0$ that satisfies (36).

Proof. The choice of λ implies that $\partial_k C_\beta^{(k)}(0;\lambda) = 0$. Hence strategy $(f^{(k)}, g^*)$ is λ-optimal.

Note that, (36) can also be written as $\lambda = \big((M_\beta^{(k)}(0))^2 \partial_k D_\beta^{(k)}(0)\big)/\partial_k M_\beta^{(k)}(0)$. By Lemma 4.2, $\partial_k M_\beta^{(k)}(0) > 0$ and by Lemma 4.3, $\partial_k D_\beta^{(k)}(0) \geq 0$. Hence, for any $k > 0$, λ given by (36) is positive. This completes the proof. □

6.2 Result for Problem 2

By Proposition 4.4, for $\beta \in (0,1]$, the distortion-transmission function $D_\beta^*(\alpha)$ is continuous and decreasing function of α. It can be completely characterized as follows:

Theorem 7 *For any $\alpha \in (0,1)$, there exists a $k_\beta^*(\alpha)$ such that*

$$N_\beta^{(k_\beta^*(\alpha))}(0) = \alpha. \tag{37}$$

The strategy $(f^{(k_\beta^*(\alpha))}, g^*)$ is optimal for Problem 2 when $\beta \in (0,1]$. Moreover, the distortion-transmission function $D_\beta^*(\alpha)$ is given by

$$D_\beta^*(\alpha) = D_\beta^{(k_\beta^*(\alpha))}(0). \tag{38}$$

Proof. A strategy (f°, g°) is optimal for a constrained optimization problem if the following conditions hold [37]:

(C1) $N(f^\circ, g^\circ) = \alpha$,
(C2) There exists a $\lambda^\circ \geq 0$ such that (f°, g°) is optimal for $C(f, g; \lambda^\circ)$.

We will show that for a given α, there exists a $k_\beta^*(\alpha) \in \mathbb{R}_{\geq 0}$ such that $(f^{(k_\beta^*(\alpha))}, g^*)$ satisfy conditions (C1) and (C2).

By Corollary 6.1, $N_\beta^{(k)}(0)$ is continuous and strictly decreasing in k. It is easy to see that $\lim_{k\to 0} N_\beta^{(k)}(0) = 1$ and $\lim_{k\to\infty} N_\beta^{(k)}(0) = 0$. Hence, for a given $\alpha \in (0,1)$, there exists a $k_\beta^*(\alpha)$ such that $N_\beta^{(k_\beta^*(\alpha))}(0) = N_\beta(f^{(k_\beta^*(\alpha))}, g^*) = \alpha$. Thus, $(f^{(k_\beta^*(\alpha))}, g^*)$ satisfies (C1).

Now, for $k_\beta^*(\alpha)$, we can find a λ satisfying (36) and hence we have by Theorem 6 that $(f^{(k_\beta^*(\alpha))}, g^*)$ is optimal for $C_\beta(f, g; \lambda)$. Thus, $(f^{(k_\beta^*(\alpha))}, g^*)$ satisfies (C2). Hence, $(f^{(k_\beta^*(\alpha))}, g^*)$ is optimal for Problem 2.

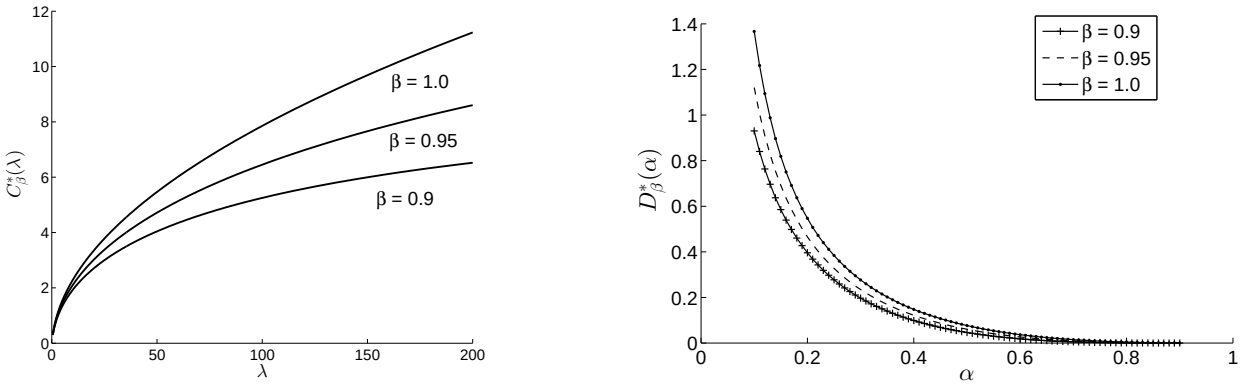
Lastly, the optimal distortion, namely the *distortion-transmission function*, which is function of α, is given by $D_\beta^*(\alpha) \coloneqq D_\beta(f^{(k_\beta^*(\alpha))}, g^*) = D_\beta^{(k_\beta^*(\alpha))}(0)$. This completes the proof. □

6.3 Computation of $C_\beta^*(\lambda)$ and $D_\beta^*(\alpha)$ for $\sigma^2 = 1$

As discussed in Section 4.3, $L_\beta^{(k)}(e)$ and $M_\beta^{(k)}(e)$ can be computed by numerically solving the Fredholm integral equations (17) and (18). We use the Matlab implementation presented in [33]. Using the result of Proposition 4.3, we can use $L_\beta^{(k)}(e)$ and $M_\beta^{(k)}(e)$ to compute $D_\beta^{(k)}(e)$ and $N_\beta^{(k)}(e)$, as well as to numerically compute $\partial_k D_\beta^{(k)}(e)$ and $\partial_k N_\beta^{(k)}(e)$. These can be combined with a binary search to compute $C_\beta^*(\lambda)$ and $D_\beta^*(\alpha)$, as shown in Algorithms 1 and 2. Fig. 4 shows the plot of $C_\beta^*(\lambda)$ and $D_\beta^*(\alpha)$ for $\sigma^2 = 1$.

6.4 Scaling with variance for Model B

In this section, we investigate the scaling of the distortion-transmission function with the variance σ^2 of the increments W_t. To show the dependence on σ, we remove the subscript β and parameterize $L_\beta^{(k)}$, $M_\beta^{(k)}$, $D_\beta^{(k)}$, $N_\beta^{(k)}$, $\mathcal{B}^{(k)}$, k_β^*, C_β^* and D^* by subscript σ.



(a) (b)

Figure 4: In Model B for $\sigma^2 = 1$, (a) $C^*_\beta(\lambda)$, and (b) $D^*_\beta(\alpha)$.

Algorithm 1: Compute transition matrices

input : $\lambda \in \mathbb{R}_{>0}$, $\beta \in (0, 1]$, $\varepsilon \in \mathbb{R}_{>0}$
output: $C^{(k^\circ)}_\beta(\lambda)$, where $|k^\circ - k^*_\beta(\lambda)| < \varepsilon$
Let $\lambda^*_\beta(k)$ denote the left-hand side of (36)
Pick $\underline{k}$ and $\bar{k}$ such that $\lambda^*_\beta(\underline{k}) < \lambda < \lambda^*_\beta(\bar{k})$
$k^\circ = (\underline{k} + \bar{k})/2$
while $|\lambda^*_\beta(k^\circ) - \lambda| > \varepsilon$ **do**
  if $\lambda^*(k^\circ) < \lambda$ **then**
    $\underline{k} = k^\circ$
  else
    $\bar{k} = k^\circ$
  $k^\circ = (\underline{k} + \bar{k})/2$
return $D^{(k^\circ)}_\beta(0) + \lambda N^{(k^\circ)}_\beta(0)$

Algorithm 2: Compute transition matrices

input : $\alpha \in (0, 1)$, $\beta \in (0, 1]$, $\varepsilon \in \mathbb{R}_{>0}$
output: $D^{(k^\circ)}_\beta(\alpha)$, where $|N^{(k^\circ)}_\beta(0) - \alpha| < \varepsilon$
Pick $\underline{k}$ and $\bar{k}$ such that $N^{(\underline{k})}_\beta(0) < \alpha < N^{(\bar{k})}_\beta(0)$
$k^\circ = (\underline{k} + \bar{k})/2$
while $|N^{(k^\circ)}_\beta(0) - \alpha| > \varepsilon$ **do**
  if $N^{(k^\circ)}_\beta(0) < \alpha$ **then**
    $\underline{k} = k^\circ$
  else
    $\bar{k} = k^\circ$
  $k^\circ = (\underline{k} + \bar{k})/2$
return $D^{(k^\circ)}_\beta(\alpha)$

Lemma 6.1 *For Model B, let $L_\sigma^{(k)}$ and $M_\sigma^{(k)}$ be the solutions of (17) and (18) respectively, when the variance of W_t is σ^2. Then*

$$L_\sigma^{(k)}(e) = \sigma^2 L_1^{(k/\sigma)}\Big(\frac{e}{\sigma}\Big), \quad M_\sigma^{(k)}(e) = M_1^{(k/\sigma)}\Big(\frac{e}{\sigma}\Big), \tag{39}$$

$$D_\sigma^{(k)}(e) = \sigma^2 D_1^{(k/\sigma)}\Big(\frac{e}{\sigma}\Big), \quad N_\sigma^{(k)}(e) = N_1^{(k/\sigma)}\Big(\frac{e}{\sigma}\Big). \tag{40}$$

Proof. For any $\beta \in (0,1]$, $L_\sigma^{(k)}$ is the solution of the following equation

$$\Big[L_\sigma^{(k)} - \beta \mathcal{B}_\sigma^{(k)} L_\sigma^{(k)}\Big](e) = e^2.$$

Define $\hat{L}_\sigma^{(k)}(e) \coloneqq \sigma^2 L_1^{(k/\sigma)}\Big(\frac{e}{\sigma}\Big)$. Then, it can be shown using first principles that

$$\beta\Big[\mathcal{B}_\sigma^{(k)} \hat{L}_\sigma^{(k)}\Big](e) = \beta\sigma^2\Big[\mathcal{B}_1^{(k/\sigma)} L_1^{(k/\sigma)}\Big]\Big(\frac{e}{\sigma}\Big). \tag{41}$$

Therefore,

$$\begin{aligned}\Big[\hat{L}_\sigma^{(k)} - \beta\mathcal{B}_\sigma^{(k)}\hat{L}_\sigma^{(k)}\Big](e) &= \sigma^2\Big[L_1^{(k/\sigma)} - \beta\mathcal{B}_1^{(k/\sigma)} L_1^{(k/\sigma)}\Big]\Big(\frac{e}{\sigma}\Big)\\ &= \sigma^2\frac{e^2}{\sigma^2} = e^2.\end{aligned}$$

This proves the scaling of $L_\sigma^{(k)}$. The scaling of $M_\sigma^{(k)}$ can be proved similarly. The scaling of $D_\sigma^{(k)}$ and $N_\sigma^{(k)}$ follow from Proposition 4.3. This completes the proof. □

Theorem 8 *For Problem 1, $k_\sigma^*(\lambda) = k_1^*(\lambda/\sigma^2)$ and $C_\sigma^*(\lambda) = \sigma^2 C_1^*(\lambda/\sigma^2)$.*

Proof. By definition of total communication cost, we have that

$$\begin{aligned}C_\sigma^{(k)}(0;\lambda) &= D_\sigma^{(k)}(0) + \lambda N_\sigma^{(k)}(0)\\ &\overset{(a)}{=} \sigma^2 D_1^{(k)}(0) + \lambda N_1^{(k)}(0)\\ &= \sigma^2 C_1^{(k)}(0;\frac{\lambda}{\sigma^2}),\end{aligned} \tag{42}$$

where the equality (a) follows from Lemma 6.1. Since $k_\sigma^*(\lambda) = \arg\min_{k\in\mathbb{R}_{\geq 0}} C_\sigma^{(k)}(0;\lambda)$ and $C_\sigma^*(\lambda) = C_\sigma^{(k_\sigma^*(\lambda))}(\lambda)$, the proof follows from (42). □

Theorem 9 *For Problem 2, $k_\sigma^*(\alpha) = \sigma k_1^*(\alpha)$ and $D_\sigma^*(\alpha) = \sigma^2 D_1^*(\alpha)$.*

Proof. The scaling of $k^*(\alpha)$ follows from the definition in Proposition 7 and the scaling properties shown in Lemma 6.1. Now,

$$\begin{aligned}D_\sigma^*(\alpha) = D_\sigma^{(k_\sigma^*(\alpha))}(0) &= D_\sigma^{(\sigma k_1^*(\alpha))}(0)\\ &\overset{(a)}{=} \sigma^2 D_1^{(k_1^*(\alpha))}(0) = \sigma^2 D_1^*(\alpha),\end{aligned}$$

where equality (a) is obtained by using (40). □

An implication of the above theorem is that we only need to numerically compute $C_1^*(\lambda)$ and $D_1^*(\alpha)$. The optimal total communication cost and the distortion-transmission function for any other value of σ^2 can be obtained by simply scaling $C_1^*(\lambda)$ and $D_1^*(\alpha)$ respectively.

7 An example for Model A

An example of a source and a distortion function that satisfy Model A is the following:

Example 1 *Consider a Markov chain of the form (1) where the pmf of W_t is given by*

$$p_n = \begin{cases} p, & \text{if } |n| = 1 \\ 1-2p, & \text{if } n = 0 \\ 0, & \text{otherwise,} \end{cases}$$

where $p \in (0, \frac{1}{3})$. The distortion function is taken as $d(e) = |e|$.

Note that the Markov process corresponds to a symmetric, birth-death Markov chain defined over $\mathbb{Z}$ as shown in Fig. 5, with the transition probability matrix is given by

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

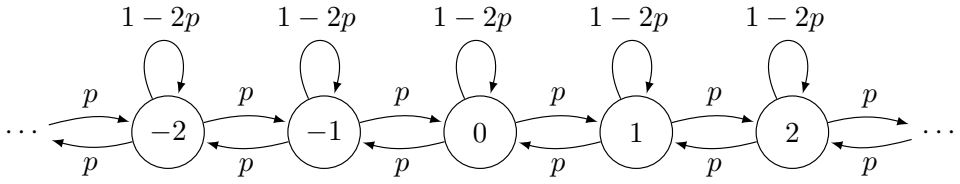


Figure 5: A birth-death Markov chain

Remark 7.1 The model of Example 1 satisfies (A1) with $\rho(e) = \max\{\lambda, |e|\}$, $\mu_1 = 1$, and $\mu_2 = \max\{1-2p+2p/\lambda, 2\}$. This may be verified by direct substitution.

In this section, we characterize $C^*_\beta(\alpha)$ and $D^*_\beta(\alpha)$ for the birth-death Markov chain presented in Example 1. As shown in Remark 7.1, this model satisfies Assumption (A1). Thus, we can use (30) to compute $\{\lambda^{(k)}_\beta\}_{k=0}^\infty$. The result of Theorem 3 is given in terms of $L^{(k)}_\beta$ and $M^{(k)}_\beta$, which, in turn, depend on the matrix $Q^{(k)}_\beta$. The matrix $Q^{(k)}_\beta$ is the inverse of a tridiagonal symmetric Toeplitz matrix and an explicit formula for its elements is available [42].

Lemma 7.1 *Define for $\beta \in (0, 1]$*

$$K_\beta = -2 - \frac{(1-\beta)}{\beta p} \quad \textit{and} \quad m_\beta = \cosh^{-1}(-K_\beta/2)$$

Then,

$$[Q^{(k)}_\beta]_{ij} = \frac{1}{\beta p} \frac{[A^{(k)}_\beta]_{ij}}{b^{(k)}_\beta}, \quad i, j \in S^{(k)},$$

where, for $\beta \in (0, 1)$,

$$\begin{aligned} [A^{(k)}_\beta]_{ij} &= \cosh((2k - |i-j|)m_\beta) - \cosh((i+j)m_\beta), \\ b^{(k)}_\beta &= \sinh(m_\beta)\sinh(2km_\beta); \end{aligned}$$

and for $\beta = 1$,

$$[A_1^{(k)}]_{ij} = (k - \max\{i,j\})(k + \min\{i,j\}),$$
$$b_1^{(k)} = 2k.$$

In particular, the elements $[Q_\beta^{(k)}]_{0j}$ *are given as follows. For* $\beta \in (0,1)$,

$$[Q_\beta^{(k)}]_{0j} = \frac{1}{\beta p} \frac{\cosh((2k - |j|)m_\beta) - \cosh(jm_\beta)}{2\sinh(m_\beta)\sinh(2km_\beta)}, \tag{43}$$

and for $\beta = 1$,

$$[Q_1^{(k)}]_{0j} = \frac{k - |j|}{2p}. \tag{44}$$

Proof. The matrix $I_{2k-1} - \beta P^{(k)}$ is a symmetric tridiagonal matrix given by

$$I_{2k-1} - \beta P^{(k)} = -\beta p \begin{bmatrix} K_\beta & 1 & 0 & \cdots & \cdots & 0 \\ 1 & K_\beta & 1 & 0 & \cdots & 0 \\ 0 & 1 & K_\beta & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & K_\beta & 1 \\ 0 & 0 & \cdots & 0 & 1 & K_\beta \end{bmatrix}.$$

$Q_\beta^{(k)}$ is the inverse of the above matrix. The inverse of the tridiagonal matrix in the above form with $K_\beta \leq -2$ are computed in closed form in [42]. The result of the lemma follows from these results. □

Using the expressions for $Q_\beta^{(k)}$, we obtain closed form expressions for $L_\beta^{(k)}$ and $M_\beta^{(k)}$.

Lemma 7.2

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

$$D_\beta^{(k)}(0) = \frac{\sinh(km_\beta) - k\sinh(m_\beta)}{2\sinh^2(km_\beta/2)\sinh(m_\beta)};$$
$$N_\beta^{(k)}(0) = \frac{2\beta p \sinh^2(m_\beta/2)\cosh(km_\beta)}{\sinh^2(km_\beta/2)} - (1-\beta).$$

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

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

and

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

Proof. By substituting the expression for $Q_\beta^{(k)}$ from Lemma 7.1 in the expressions for $L_\beta^{(k)}$ and $M_\beta^{(k)}$ from Proposition 4.1, we get that

1. For $\beta \in (0,1)$,

$$L_\beta^{(k)}(0) = \frac{\sinh(km_\beta) - k\sinh(m_\beta)}{4\beta p \sinh^2(m_\beta/2)\sinh(m_\beta)\cosh(km_\beta)},$$
$$M_\beta^{(k)}(0) = \frac{\sinh^2(km_\beta/2)}{2\beta p \sinh^2(m_\beta/2)\cosh(km_\beta)}.$$

2. For $\beta = 1$,

$$L_1^{(k)}(0) = k(k^2-1)/(6p), \quad M_1^{(k)}(0) = k^2/(2p).$$

The results of the lemma follow using the above expressions and Proposition 4.3. The expression for $\lambda_1^{(k)}$ is obtained by plugging the expressions of $D_1^{(k+1)}$, $D_1^{(k)}$, $N_1^{(k+1)}$, and $N_1^{(k)}$ in (30). □

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 1.

For $\beta = 1$, we can use the analytic expression of $\lambda^{(k)}$ to verify that $\{\lambda_\beta^{(k)}\}_{k=0}^{\infty}$ is increasing. For $\beta \in (0,1)$, we can numerically verify that $\{\lambda_\beta^{(k)}\}_{k=0}^{\infty}$ is increasing. Thus we can use the result of Theorem 3 to compute $C_\beta^*(\lambda)$. See Fig. 6 for the plot of $C_\beta^*(\lambda)$ vs λ for different values of β (all for $p = 0.3$). An alternative way to plot this curve is to draw the vertices $(D_\beta^{(k)}(0) + \lambda N_\beta^{(k)}(0), \lambda_\beta^{(k)})$ using the data in Table 1 and join any pair of vertices with a straight line. The optimal total communication cost for a given λ can then be found from the data.

For example, for $\lambda = 20$, $\beta = 0.9$, we can find from Table 1a that $\lambda \in (\lambda_\beta^{(4)}, \lambda_\beta^{(5)}]$. Hence, $k_\beta^* = 5$ (i.e. the strategy $f^{(5)}$ is optimal) and the optimal total communication cost is computed from the table as

$$C_{0.9}^*(20) = D_{0.9}^{(5)}(0) + 20N_{0.9}^{(5)}(0) = 1.1844 + 20 \times 0.0111 = 1.4064.$$

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

(a) For $\beta = 0.9$

k	$D_\beta^{(k)}(0)$	$N_\beta^{(k)}(0)$	$\lambda_\beta^{(k)}$
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	$D_\beta^{(k)}(0)$	$N_\beta^{(k)}(0)$	$\lambda_\beta^{(k)}$
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	$D_\beta^{(k)}(0)$	$N_\beta^{(k)}(0)$	$\lambda_\beta^{(k)}$
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

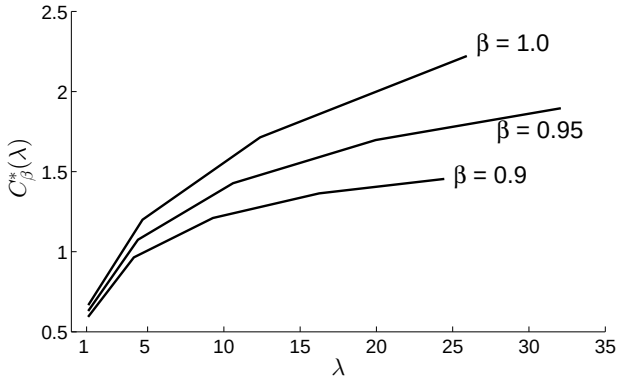


Figure 6: Plot of $C_\beta^*(\lambda)$ vs λ for the Markov chain of Example 1 with $p = 0.3$.

Lemma 7.3 *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

$$k^*_1 = \left\lfloor \sqrt{\frac{2p}{\alpha}} \right\rfloor.$$

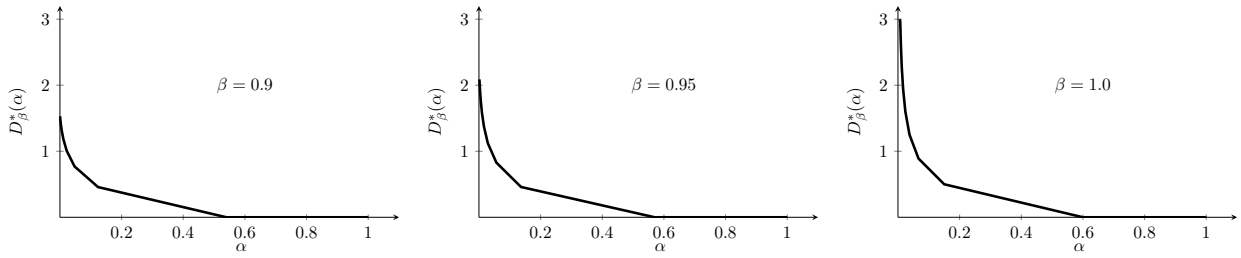
Proof. The result of the lemma follows directly by using the definition of k^*_β given in (31) in the expressions given in Lemma 7.2. □

Using the above results, we can plot and the distortion-transmission function $D^*_\beta(\alpha)$. See Fig. 7 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 $(N^{(k)}_\beta, D^{(k)}_\beta)$ using the data in Table 1 to compute the optimal (randomized) strategy for a particular value of α.

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 $N^{(k)}_\beta \geq \alpha$. Thus, from Table 1a, we get that $k^* = 1$. Then, by (33),

$$\theta^* = \frac{\alpha - N^{(2)}_\beta}{N^{(1)}_\beta - N^{(2)}_\beta} = 0.9039.$$

Let $f^* = (f^{(1)}, f^{(2)}, \theta^*)$. Then the Bernoulli randomized simple strategy (f^*, g^*) is optimal for Problem 2 for $\beta \in (0,1)$. Furthermore, by (35), $D^*_\beta(\alpha) = 0.044$.



(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 $\beta = 1.0$

Figure 7: Plots of $D^*_\beta(\alpha)$ vs α for different β for the birth-death Markov chain of Example 1 with $p = 0.3$.

8 Conclusion

We characterize two fundamental limits of remote estimation of Markov processes under communication constraints. First, when each transmission is costly, we characterize the minimum achievable cost of communication plus estimation error. Second, when there is a constraint on the average number of transmissions, we characterize the minimum achievable estimation error.

We also identify transmission and estimation strategies that achieve these fundamental limits. The structure of these optimal strategies had been previously identified by using dynamic programming for decentralized stochastic control systems. In particular, the optimal transmission strategy is to transmit when the estimation error process exceeds a threshold and the optimal estimation strategy is to select the last transmitted state as the estimate. We identify the performance of a generic strategy that has such a structure.

For the case of costly communication, we identify the value of communication cost for which a particular threshold-based strategy is optimal; for the case of constrained communication, we identify (possibly randomized) threshold-based strategies that achieve the communication constraint.

The results are derived under an idealized system model. In particular, we assume that when the transmitter does transmit, it sends the complete state of the source; the channel is noiseless and does not introduce any delay. Relaxing these assumptions to analyze the effects of quantization, channel noise and delay are important future directions.

Appendix A Proof of Proposition 3.1

We only prove the result for Model A. The proof for Model B is identical. To prove this proposition, we first consider the finite horizon setup and show that the value functions are even and increasing. The result of Proposition 3.1 follows, because monotonicity is preserved under limits.

Definition A.1 (Stochastic Dominance) *Let μ and ν be two probability distributions defined over $\mathbb{Z}_{\geq 0}$. Then μ is said to dominate ν in the sense of stochastic dominance, which is denoted by $\mu \succeq_s \nu$, if*

$$\sum_{i \geq n} \mu_i \geq \sum_{i \geq n} \nu_i, \quad \forall n \in \mathbb{Z}_{\geq 0}.$$

A very useful property of stochastic dominance is the following:

Lemma A.1 *For any probability distributions μ and ν on $\mathbb{Z}_{\geq 0}$ such that $\mu \succeq_s \nu$ and for any increasing function $f : \mathbb{Z}_{\geq 0} \to \mathbb{R}$,*

$$\sum_{n=0}^{\infty} f(n)\mu_n \geq \sum_{n=0}^{\infty} f(n)\nu_n.$$

This is a standard result. See, for example, [29, Lemma 4.7.2].

To prove Proposition 3.1, we extend the notion of stochastic dominance to distributions defined over $\mathbb{Z}$.

Definition A.2 (Reflected stochastic dominance) *Let μ and ν be two probability distributions defined over $\mathbb{Z}$. Then μ is said to dominate ν in the sense of reflected stochastic dominance, which is denoted by $\mu \succeq_r \nu$, if*

$$\sum_{i \geq n} (\mu_i + \mu_{-i}) \geq \sum_{i \geq n} (\nu_i + \nu_{-i}), \quad \forall n \in \mathbb{Z}_{>0}.$$

Lemma A.2 *For any probability distributions μ and ν defined over $\mathbb{Z}$ such that $\mu \succeq_r \nu$ and for any function $f : \mathbb{Z} \to \mathbb{R}$ that is even and increasing on $\mathbb{Z}_{\geq 0}$,*

$$\sum_{n=-\infty}^{\infty} f(n)\mu_n \geq \sum_{n=-\infty}^{\infty} f(n)\nu_n.$$

Proof. Define distributions $\tilde{\mu}$ and $\tilde{\nu}$ over $\mathbb{Z}_{\geq 0}$ as follows: for every $n \in \mathbb{Z}_{\geq 0}$

$$\tilde{\mu}_n = \begin{cases} \mu_0, & \text{if } n = 0 \\ \mu_n + \mu_{-n}, & \text{otherwise;} \end{cases}$$

and $\tilde{\nu}$ defined similarly.

By definition, $\mu \succeq_r \nu$ implies that $\tilde{\mu} \succ_s \tilde{\nu}$; hence, the result follows from Lemma A.1. □

Lemma A.3 *Define a sequence of probability distributions $\{\mu_e : e \in \mathbb{Z}\}$ as follows: for any $n \in \mathbb{Z}$, $\mu_{e,n} = p_{e+n}$. Then $\mu_{e+1} \succeq_r \mu_e$.*

Proof. To prove the result, we have to show that for any $n \in \mathbb{Z}_{\geq 0}$

$$\sum_{i \geq n+1} (p_{i-e-1} + p_{-i-e-1}) \geq \sum_{i \geq n+1} (p_{i-e} + p_{-i-e}),$$

or, equivalently,

$$\sum_{i=-n}^{n} p_{i-e} \geq \sum_{i=-n}^{n} p_{i-e-1}.$$

To prove the above, it is sufficient to show that

$$p_{i-e} \geq p_{-i-e-1}, \quad \forall e, i \in \mathbb{Z}_{\geq 0}. \tag{45}$$

Recall that $\{p_n\}_{n=0}^{\infty}$ is a decreasing sequence. Since e and i are positive, we have that $i-e \leq i+e < i+e+1$. Hence, $p_{i-e} \geq p_{i+e+1} = p_{-i-e-1}$, which proves (45). □

Finally, note the following obvious properties of even and increasing functions that we state without proof. Let EI denote 'even and increasing on $\mathbb{Z}_{\geq 0}$'. Then

(P1) Sum of two EI functions is EI.

(P2) Pointwise minimum of two EI functions is EI.

We now prove Proposition 3.1 for Model A.

Proof of Proposition 3.1. We prove the result by backward induction. The result is trivially true for V_T, which is the basis of induction. Assume that $V_{t+1}(\cdot;\lambda)$ is even and increasing on $\mathbb{Z}_{\geq 0}$. Define

$$\hat{V}_t(e;\lambda) = \sum_{n=-\infty}^{\infty} \mu_{e,-n} V_{t+1}(n;\lambda),$$

where $\mu_{e,n}$ is defined in Lemma A.3. We show that $\hat{V}_t(\cdot;\lambda)$ is even and increasing on $\mathbb{Z}_{\geq 0}$.

1. Consider

$$\begin{aligned}\hat{V}_t(-e;\lambda) &= \sum_{n=-\infty}^{\infty} \mu_{-e,-n} V_{t+1}(n;\lambda) \\ &= \sum_{-n=-\infty}^{\infty} p_{-e+n} V_{t+1}(-n;\lambda) \\ &\overset{(a)}{=} \sum_{n=-\infty}^{\infty} p_{e-n} V_{t+1}(n;\lambda) = \hat{V}_t(e;\lambda)\end{aligned}$$

   where (a) uses $p_w = p_{-w}$ and $V_{t+1}(e+w;\lambda) = V_{t+1}(-e-w;\lambda)$. Hence, $\hat{V}_t(\cdot;\lambda)$ is even.

2. By Lemma A.3, for all $e \in \mathbb{Z}_{\geq 0}$, $\mu_{e+1} \succeq_r \mu_e$. Since $V_{t+1}(\cdot;\lambda)$ is even and increasing on $\mathbb{Z}_{\geq 0}$, by Lemma A.2, $\hat{V}_t(e+1;\lambda) \geq \hat{V}_t(e;\lambda)$. Hence, $\hat{V}_t(\cdot;\lambda)$ is increasing on $\mathbb{Z}_{\geq 0}$.

Now, V_t is given by

$$V_t(e;\lambda) = \min\left\{\lambda + \hat{V}_t(0;\lambda), d(e) + \hat{V}_t(e;\lambda)\right\}.$$

Since $\lambda_\beta^{(k)}$ is increasing in k, $d(\cdot)$ is even and increasing on $\mathbb{Z}_{\geq 0}$. Therefore, by properties (P1) and (P2) given above, the function $V_t(\cdot;\lambda)$ is even and increasing on $\mathbb{Z}_{\geq 0}$. This completes the induction step. Therefore, the result of Proposition 3.1 follows from the principle of induction. □

Appendix B Proof of Proposition 4.3

We prove the result for the discounted cost setup, $\beta \in (0,1)$. The result extends to the long-term average cost setup, $\beta = 1$ by using the vanishing discount approach similar to the argument given in Section 3.3.

We first consider the case $k = 0$. In this case, the recursive definition of $D_\beta^{(k)}$ and $N_\beta^{(k)}$, given by (26) and (27), simplify to the following:

$$D_\beta^{(0)}(e) = \beta[\mathcal{B}D_\beta^{(0)}](0);$$

and

$$N_\beta^{(0)}(e) = (1-\beta) + \beta[\mathcal{B}N_\beta^{(0)}](0).$$

It can be easily verified that $D_\beta^{(0)}(e) = 0$ and $N_\beta^{(0)}(e) = 1$, $e \in \mathbb{Z}$ for Model A and $e \in \mathbb{R}$ Model B, satisfy the above equations. Also, $C_\beta^{(0)}(e;\lambda) = C_\beta(f^{(0)}, g^*; \lambda) = \lambda$. This proves the first part of the proposition.

For $k > 0$, let $\tau^{(k)}$ denote the stopping time when the Markov process in both Model A and B starting at state 0 at time $t = 0$ enters the set $S^{(k)}$. Note that $\tau^{(0)} = 1$ and $\tau^{(\infty)} = \infty$.

Then,

$$L_\beta^{(k)}(0) = \mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_0) \;\Big|\; E_0 = 0\Big] \tag{46}$$

$$\begin{aligned} M_\beta^{(k)}(0) &= \mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t \;\Big|\; E_0 = 0\Big] \\ &= \frac{1 - \mathbb{E}[\beta^{\tau^{(k)}} \,|\, E_0 = 0]}{1-\beta} \end{aligned} \tag{47}$$

$$\begin{aligned} D_\beta^{(k)}(0) = \mathbb{E}\Big[&(1-\beta)\sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_0) \\ &+ \beta^{\tau^{(k)}} D_\beta^{(k)}(0) \;\Big|\; E_t = 0\Big] \end{aligned} \tag{48}$$

$$\begin{aligned} N_\beta^{(k)}(0) = \mathbb{E}\Big[&(1-\beta)\sum_{t=0}^{\tau^{(k)}-1} \beta^t \\ &\beta^{\tau^{(k)}} N_\beta^{(k)}(0) \;\Big|\; E_t = 0\Big]. \end{aligned} \tag{49}$$

Substituting (46) and (47) in (48) we get

$$D_\beta^{(k)}(0) = (1-\beta)L_\beta^{(k)}(0) + [1 - (1-\beta)M_\beta^{(k)}(0)]D_\beta^{(k)}(0).$$

Rearranging, we get that

$$D_\beta^{(k)}(0) = \frac{L_\beta^{(k)}(0)}{M_\beta^{(k)}(0)}.$$

Similarly, substituting (46) and (47) in (49) we get

$$N_\beta^{(k)}(0) = [1 - (1-\beta)M_\beta^{(k)}(0)][(1-\beta) + N_\beta^{(k)}(0)].$$

Rearranging, we get that

$$N_\beta^{(k)}(0) = \frac{1}{M_\beta^{(k)}(0)} - (1-\beta).$$

The expression for $C_\beta^{(k)}(0;\lambda)$ follows from the definition.

Appendix C Proof of Lemma 4.3

We prove the strict monotonicity of $D_\beta^{(k)}$ in k for Model A for $\beta \in (0,1)$. The result for $\beta = 1$ follows by taking limit $\beta \uparrow 1$. The result for Model B is similar. To prove the result, we show that for any $k \in \mathbb{Z}_{\geq 0}$, $D_\beta^{(k)}(e) < D_\beta^{(k+1)}(e)$, $\forall e \in \mathbb{Z}$.

For any $\beta \in (0,1)$ and $k \in \mathbb{Z}_{\geq 0}$, define the operator $T^{(k)} : (\mathbb{Z} \to \mathbb{R}) \to (\mathbb{Z} \to \mathbb{R})$ as follows. For any $D : \mathbb{Z} \to \mathbb{R}$,

$$[\mathcal{T}^{(k)} D](e) = \begin{cases} \beta[\mathcal{B}D](0), & \text{if } |e| \geq k \\ (1-\beta)d(e) + \beta[\mathcal{B}D](e) & \text{if } |e| < k. \end{cases} \tag{50}$$

Define $D_\beta^{(k,0)} = D_\beta^{(k)}$, and for $m \in \mathbb{Z}_{>0}$, $D_\beta^{(k,m)} = \mathcal{T}^{(k+1)} D_\beta^{(k,m-1)}$.

Let $b \coloneqq \sup\{k \in \mathbb{Z}_{\geq 0} \,|\, p_k > 0\}$ and define $A^{(m)} = A_+^{(m)} \cup A_-^{(m)}$, where

$$\begin{aligned} A_+^{(m)} &= \{k, k-1, \cdots, \max(k-mb, 0)\}, \\ A_-^{(m)} &= \{-k, -k+1, \cdots, \min(-k+mb, 0)\} \end{aligned}$$

Note that $A^{(0)} = \{-k, k\}$ and $A^{(m)} \subseteq A^{(m+1)} \subseteq \{-k, \cdots, k\}$. Let m° be the smallest integer such that $A^{(m^\circ)} = \{-k, \cdots, k\}$ (in particular, if $b = \infty$, then $m^\circ = 2$).

Next, define $B^{(0)} = \emptyset$ and for $m \in \mathbb{Z}_{>0}$ $B^{(m)} = B_+^{(m)} \cup B_-^{(m)}$, where

$$\begin{aligned} B_+^{(m)} &= \{k+1, \cdots, k+mb\}, \\ B_-^{(m)} &= \{-k-1, \cdots, -k-mb\} \end{aligned}$$

See Appendices C and D of the supplementary document for the proofs of the following two results.

Lemma C.1 *For any* $m \in \{0, 1, \cdots, m^\circ\}$

$$D_\beta^{(k,m+1)}(e) > D_\beta^{(k)}(e), \quad \forall e \in A^{(m)}$$

and

$$D_\beta^{(k,m+1)}(e) \geq D_\beta^{(k)}(e), \quad \forall e \notin A^{(m)}.$$

Lemma C.2 *For* $m \in \mathbb{Z}_{\geq 0}$,

$$D_\beta^{(k,m+m^\circ+1)}(e) > D_\beta^{(k)}(e), \quad \forall e \in B^{(m)} \cup A^{(m^\circ)}$$

and

$$D_\beta^{(k,m+m^\circ+1)}(e) \geq D_\beta^{(k)}(e), \quad \forall e \notin B^{(m)} \cup A^{(m^\circ)}.$$

Since $\lim_{m\to\infty} B^{(m)} \cup A^{(m^\circ)} = \mathbb{Z}$, Lemma C.2 implies that

$$D_\beta^{(k)}(e) < \lim_{m\to\infty} D_\beta^{(k,m)}(e). \tag{51}$$

From Theorem 2, the operator $\mathcal{T}^{(k+1)}$ is a contraction and $D_\beta^{(k+1)}$ is its a unique bounded fixed point. Hence, the right hand side of (51) equals $D_\beta^{(k+1)}(e)$. This completes the proof for Model A.

References

[1] T. Linder and G. Lugosi, A zero-delay sequential scheme for lossy coding of individual sequences, IEEE Trans. Inf. Theory, 47(6), 2533–2538, 2001.

[2] T. Weissman and N. Merhav, On limited-delay lossy coding and filtering of individual sequences, IEEE Trans. Inf. Theory, 48(3), 721–733, 2002.

[3] A. György, T. Linder, and G. Lugosi, Efficient adaptive algorithms and minimax bounds for zero-delay lossy source coding, IEEE Trans. Signal Process., 52(8), 2337–2347, 2004.

[4] S. Matloub and T. Weissman, Universal zero-delay joint source-channel coding, IEEE Trans. Inf. Theory, 52(12), 5240–5250, Dec. 2006.

[5] H.S. Witsenhausen, On the structure of real-time source coders, Bell System Technical Journal, 58(6), 1437–1451, July-August 1979.

[6] J.C. Walrand and P. Varaiya, Optimal causal coding-decoding problems, IEEE Trans. Inf. Theory, 29(6), 814–820, Nov. 1983.

[7] D. Teneketzis, On the structure of optimal real-time encoders and decoders in noisy communication, IEEE Trans. Inf. Theory, 4017–4035, Sep. 2006.

[8] A. Mahajan and D. Teneketzis, Optimal design of sequential real-time communication systems, IEEE Trans. Inf. Theory, 55(11), 5317–5338, Nov. 2009.

[9] Y. Kaspi and N. Merhav, Structure theorems for real-time variable rate coding with and without side information, IEEE Trans. Inf. Theory, 58(12), 7135–7153, 2012.

[10] H. Asnani and T. Weissman, Real-time coding with limited lookahead, IEEE Trans. Inf. Theory, 59(6), 3582–3606, 2013.

[11] O.C. Imer and T. Basar, Optimal estimation with limited measurements, Joint 44the IEEE Conference on Decision and Control and European Control Conference, 29, 1029–1034, 2005.

[12] Y. Xu and J.P. Hespanha, Optimal communication logics in networked control systems, in Proceedings of 43rd IEEE Conference on Decision and Control, 4, 2004, 3527–3532.

[13] G.M. Lipsa and N. Martins, Remote state estimation with communication costs for first-order LTI systems, IEEE Trans. Autom. Control, 56(9), 2013–2025, Sep. 2011.

[14] 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, 58(9), 2246–2260, 2013.

[15] A. Molin and S. Hirche, An iterative algorithm for optimal event-triggered estimation, in 4th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS'12), 2012, 64–69.

[16] 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, 32(2), 554–568, April 1996.

[17] S. Appadwedula, V.V. Veeravalli, and D.L. Jones, Decentralized detection with censoring sensors, IEEE Transactions on Signal Processing, 56(4), 1362–1373, April 2008.

[18] M. Athans, On the determination of optimal costly measurement strategies for linear stochastic systems, Automatica, 8(4), 397–412, 1972.

[19] J. Geromel, Global optimization of measurement startegies for linear stochastic systems, Automatica, 25(2), 293–300, 1989.

[20] W. Wu, A. Araposthathis, and V.V. Veeravalli, Optimal sensor querying: General Markovian and LQG models with controlled observations, IEEE Transactions on Automatic Control, 53(6), 1392–1405, 2008.

[21] D. Shuman and M. Liu, Optimal sleep scheduling for a wireless sensor network node, in Proceedings of the Asilomar Conference on Signals, Systems, and Computers, October 2006, 1337–1341.

[22] M. Sarkar and R.L. Cruz, Analysis of power managemnet for energy and delay trade-off in a WLAN, in Proceedings ofthe Conference on Information Sciences and Systems,, March 2004.

[23] M. Sarkar and R.L. Cruz, An adaptive sleep algorithm for efficient power management in WLANs, in Proceedings of the Vehicular Technology Conference, May 2005.

[24] A. Federgruen and K.C. So, Optimality ofthreshold policies in single- server queueing systems with server vacations, Adv. Appl. Prob., 23(2), 388–405, 1991.

[25] K.J. Åström, Analysis and Design of Nonlinear Control Systems. Berlin, Heidelberg: Springer, 2008, ch. Event based control.

[26] M. Rabi, G. Moustakides, and J. Baras, Adaptive sampling for linear state estimation, SIAM Journal on Control and Optimization, 50(2), 672–702, 2012.

[27] X. Meng and T. Chen, Optimal sampling and performance comparison of periodic and event based impulse control, IEEE Transactions of Automatic Control, 57(12), 3252–3259, 2012.

[28] L. Wang, J. Woo, and M. Madiman, A lower bound on Rényi entropy of convolutions in the integers, in Proceedings of the 2014 IEEE International Symposium on Information Theory, Jul. 2014, 2829–2833.

[29] M. Puterman, Markov decision processes: Discrete Stochastic Dynamic Programming. John Wiley and Sons, 1994.

[30] O.H. Lerma and J.B. Lasserre, Discrete-time Markov control processes: Basic optimality criteria, ser. Applications of mathematics 30. Springer, 1996.

[31] L.I. Sennott, Stochastic dynamic programming and the control of queueing systems. New York, NY, USA: Wiley, 1999.

[32] A.D. Polyanin and A.V. Manzhirov, Handbook of integral equations, 2nd ed. Chapman and Hall/CRC Press, 2008.

[33] K. Atkinson and L.F. Shampine, Solving Fredholm integral equations of the second kind in Matlab, ACM Trans. Math. Software, 2008.

[34] D.M. Topkis, Supermodularity and complementarity, ser. Frontiers of economic research. Princeton, NJ, USA: Princeton University Press, 1998.

[35] L.I. Sennott, Computing average optimal constrained policies in stochastic dynamic programming, Probability in the Engineering and Informational Sciences, 15, 103–133, 2001.

[36] V. Borkar, A convex analytic approach to Markov decision processes, Probability Theory and Related Fields, 78(4), 583–602, 1988.

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

[38] E. Feinberg, Optimality of deterministic policies for certain stochastic control problems with multiple criteria and constraints, in Mathematical Control Theory and Finance, A. Sarychev, A. Shiryaev, M. Guerra, and M. Grossinho, Eds. Springer Berlin Heidelberg, 2008, 137–148.

[39] A. Shwartz and A.M. Makowski, An optimal adaptive scheme for two competing queues with constraints, in Analysis and optimization of systems. Springer Berlin Heidelberg, 1986, 515–532.

[40] D.-J. Ma, A.M. Makowski, and A. Shwartz, Stochastic approximations for finite-state Markov chains, Stochastic Processes and Their Applications, 35(1), 27–45, 1990.

[41] E. Altman and A. Shwartz, Time-sharing policies for controlled Markov chains, Operations Research, 41(6), 1116–1124, 1993.

[42] G. Hu and R. O'Connell, Analytical inversion of symmetric tridiagonal matrices, Journal of Physics A: Mathematical and General, 29(7), 1511, 1996.

[43] B. Hajek, K. Mitzel, and S. Yang, Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm, IEEE Trans. Inf. Theory, 64, 608–622, Feb. 2008.

Supplementary document

Appendix A Proof of the structural results

The results of [14] relied on the notion of ASU (almost symmetric and unimodal) distributions introduced in [43].

Definition A.1 (Almost symmetric and unimodal distribution) *A probability distribution μ on $\mathbb{Z}$ is almost symmetric and unimodal (ASU) about a point $a \in \mathbb{Z}$ if for every $n \in \mathbb{Z}_{\geq 0}$,*

$$\mu_{a+n} \geq \mu_{a-n} \geq \mu_{a+n+1}.$$

A probability distribution that is ASU around 0 and even (i.e., $\mu_n = \mu_{-n}$) is called ASU and even. Note that the definition of ASU and even is equivalent to even and decreasing on $\mathbb{Z}_{\geq 0}$.

Definition A.2 (ASU Rearrangement) *The ASU rearrangement of a probability distribution μ, denoted by μ^+, is a permutation of μ such that for every $n \in \mathbb{Z}_{\geq 0}$,*

$$\mu^+_n \geq \mu^+_{-n} \geq \mu^+_{n+1}.$$

We now introduce the notion of majorization for distributions supported over $\mathbb{Z}$, as defined in [28].

Definition A.3 (Majorization) *Let μ and ν be two probability distributions defined over $\mathbb{Z}$. Then μ is said to majorize ν, which is denoted by $\mu \succeq_m \nu$, if for all $n \in \mathbb{Z}_{\geq 0}$,*

$$\sum_{i=-n}^{n} \mu^+_i \geq \sum_{i=-n}^{n} \nu^+_i,$$
$$\sum_{i=-n}^{n+1} \mu^+_i \geq \sum_{i=-n}^{n+1} \nu^+_i.$$

The structure of optimal estimator in Theorem 1 were proved in two-steps in [14]. The first step relied on the following two results.

Lemma A.1 *Let μ and ν be probability distributions with finite support defined over $\mathbb{Z}$. If μ is ASU and even and ν is ASU about a, then the convolution $\mu * \nu$ is ASU about a.*

Lemma A.2 *Let μ, ν, and ξ be probability distributions with finite support defined over $\mathbb{Z}$. If μ is ASU and even, ν is ASU, and ξ is arbitrary, then $\nu \succeq_m \xi$ implies that $\mu * \nu \succeq_m \mu * \xi$.*

These results were originally proved in [43] and were stated as Lemmas 5 and 6 in [14].

The second step (in the proof of structure of optimal estimator in Theorem 1) in [14] relied on the following result.

Lemma A.3 *Let μ be a probability distribution with finite support defined over $\mathbb{Z}$ and $f \colon \mathbb{Z} \to \mathbb{R}_{\geq 0}$. Then,*

$$\sum_{n=-\infty}^{\infty} f(n)\mu_n \leq \sum_{n=-\infty}^{\infty} f^+(n)\mu^+_n.$$

We generalize the results of Lemmas A.1, A.2, and A.3 to distributions over $\mathbb{Z}$ with possibly countable support. With these generalizations, we can follow the same two step approach of [14] to prove the structure of optimal estimator as given in Theorem 1.

The structure of optimal transmitter in Theorem 1 in [14] only relied on the structure of optimal estimator. The exact same proof works in our model as well.

A.1 Generalization of Lemma A.1 to distributions supported over $\mathbb{Z}$

The proof argument is similar to that presented in [43, Lemma 6.2]. We first prove the results for $a = 0$. Assume that ν is ASU and even. For any $n \in \mathbb{Z}_{\geq 0}$, let $r^{(n)}$ denote the rectangular function from $-n$ to n, i.e.,

$$r^{(n)}(e) = \begin{cases} 1, & \text{if } |e| \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

Note that any ASU and even distribution μ may be written as a sum of rectangular functions as follows:

$$\mu = \sum_{n=0}^{\infty} (\mu_n - \mu_{n+1}) r^{(n)}.$$

It should be noted that $\mu_n - \mu_{n+1} \geq 0$ because μ is ASU and even. ν may also be written in a similar form.

The convolution of any two rectangular functions $r^{(n)}$ and $r^{(m)}$ is ASU and even. Therefore, by the distributive property of convolution, the convolution of μ and ν is also ASU and even.

The proof for the general $a \in \mathbb{Z}$ follows from the following facts:

1. Shifting a distribution is equivalent to convolution with a shifted delta function.
2. Convolution is commutative and associative.

A.2 Generalization of Lemma A.2 to distributions supported over $\mathbb{Z}$

We follow the proof idea of [28, Theorem II.1]. For any probability distribution μ, we can find distinct indices i_j, $|j| \leq n$ such that $\mu(i_j)$, $|j| \leq n$, are the $2n+1$ largest values of μ. Define

$$\mu_n(i_j) = \mu(i_j),$$

for $|j| \leq n$ and 0 otherwise. Clearly, $\mu_n \uparrow \mu$ and if μ is ASU and even, so is μ_n.

Now consider the distributions μ, ν, and ξ from Lemma A.2 but without the restriction that they have finite support. For every $n \in \mathbb{Z}_{\geq 0}$, define μ_n, ν_n, and ξ_n as above. Note that all distributions have finite support and μ_n is ASU and even and ν_n is ASU. Furthermore, since the definition of majorization remain unaffected by truncation described above, $\nu_n \succeq_m \xi_n$. Therefore, by Lemma A.2,

$$\mu_n * \nu_n \succeq_m \mu_n * \xi_n.$$

By taking limit over n and using the monotone convergence theorem, we get

$$\mu * \nu \succeq_m \mu * \xi.$$

A.3 Generalization of Lemma A.3 to distributions supported over $\mathbb{Z}$

This is an immediate consequence of [28, Theorem II.1].

Appendix B Proof of (b) of Lemma 4.2

Note that for any bounded v, $\|\mathcal{B}^{(k)}v\|_\infty$ is bounded and increasing in k. We show that $L_\beta^{(k)}(e)$ is continuous and differentiable in k. Similar argument holds for $M_\beta^{(k)}(e)$.

We show the differentiability in k. Continuity follows from the fact that differentiable functions are continuous. Note that $L_\beta^{(k)}(e)$ and $M_\beta^{(k)}(e)$ are even functions of e. Now, for any $\varepsilon > 0$ we have

$$L_\beta^{(k+\varepsilon)}(e) - L_\beta^{(k)}(e) = \beta \int_{-k}^{k} \phi(w-e)[L_\beta^{(k+\varepsilon)}(w) - L_\beta^{(k)}(w)]dw + 2\beta \int_{k}^{k+\varepsilon} \phi(w-e) L_\beta^{(k+\varepsilon)}(w)dw$$
$$= \beta \int_{-k}^{k} \phi(w-e)[L_\beta^{(k+\varepsilon)}(w) - L_\beta^{(k)}(w)]dw + 2\beta\phi(k-e) L_\beta^{(k+\varepsilon)}(k+\varepsilon)\varepsilon + O(\varepsilon^2)$$

Let $R_\beta^{(k)}(e,w)$ be the resolvent of ϕ, as given in (25). Then,

$$L_\beta^{(k+\varepsilon)}(e) - L_\beta^{(k)}(e) = 2\beta \int_{-k}^{k} R_\beta^{(k)}(e,w)\phi(k-e) L_\beta^{(k+\varepsilon)}(w)\varepsilon dw + O(\varepsilon^2)$$

This implies that

$$\left| \frac{L_\beta^{(k+\varepsilon)}(e) - L_\beta^{(k)}(e)}{\varepsilon} \right| \leq 2\beta\|\phi\|_\infty \|L_\beta^{(k)}\|_\infty \left| \int_{-k}^{k} R_\beta^{(k)}(e,w)dw \right| + O(\varepsilon).$$

Since $\mathcal{B}^{(k)}$ is a contraction, the value of the first integral in the right hand side of the above inequality is less than 1 and the result follows from the definition of differtiability.

Appendix C Proof of Lemma C.1

We show the result for Model A. The result for Model B follows similarly. We prove the result by induction. Consider $m = 0$. Analogous to Proposition 3.1, we can show that $D_\beta^{(k)}(e)$ is even and increasing in e. By Lemma A.2 and A.3, $p_{n-e} \succeq_r p_n$. Hence,

$$\sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k)}(n) \geq \sum_{n=-\infty}^{\infty} p_n D_\beta^{(k)}(n). \tag{52}$$

For $e \in A^{(0)} = \{-k, k\}$,

$$D_\beta^{(k,1)}(e) = (1-\beta)d(e) + \beta \sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k)}(n) \tag{53}$$

and

$$D_\beta^{(k)}(e) = \beta \sum_{n=-\infty}^{\infty} p_n D_\beta^{(k)}(n). \tag{54}$$

By (52), for Model A,

$$D_\beta^{(k,1)}(e) > D_\beta^{(k)}(e), \quad \forall e \in A^{(0)} \tag{55}$$
$$D_\beta^{(k,1)}(e) \overset{(a)}{=} D_\beta^{(k)}(e), \quad \forall e \notin A^{(0)}, \tag{56}$$

where the equality (a) holds since both sides have same expressions. Now, we show the result for $m = 1$. Pick any arbitrary $e \in A^{(1)}$. We have from (50)

$$D_\beta^{(k,2)}(e) = (1-\beta)d(e) + \beta \sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k,1)}(n), \tag{57}$$

Furthermore, from (26), we have

$$D_\beta^{(k)}(e) = \begin{cases} (1-\beta)d(e) + \beta \sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k)}(n), \\ \qquad\qquad e \in A^{(1)} \setminus A^{(0)} \\ \beta \sum_{n=-\infty}^{\infty} p_n D_\beta^{(k)}(n), \qquad e \in A^{(0)}. \end{cases} \tag{58}$$

Since $e \in A^{(1)}$, $p_{k-e} > 0$ and $p_{-k-e} > 0$. Hence, by (53)–(54),

$$p_{k-e} D_\beta^{(k,1)}(n) > p_{k-e} D_\beta^{(k)}(n), \quad n \in \{-k, k\}.$$

Combining the above with (55)–(56), we get

$$\sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k,1)}(n) > \sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k)}(n), \quad \forall e \in A^{(1)},$$

and hence, by (57) and (58),

$$D_\beta^{(k,2)}(e) > D_\beta^{(k)}(e), \quad \forall e \in A^{(1)} \setminus A^{(0)}. \tag{59}$$

Also, by (55) and using monotonic increasing property of $\mathcal{T}^{(k+1)}$, we get

$$D_\beta^{(k,2)}(e) \geq D_\beta^{(k,1)}(e) > D_\beta^{(k)}(e), \quad \forall e \in A^{(0)}. \tag{60}$$

Combining (59) and (60), we get that

$$D_\beta^{(k,2)}(e) > D_\beta^{(k)}(e), \quad \forall e \in A^{(1)}.$$

Furthermore, since $D_\beta^{(k,1)}(e) \geq D_\beta^{(k)}(e)$, $\forall e \in \mathbb{Z}$, by monotonic increasing property of $\mathcal{T}^{(k+1)}$,

$$D_\beta^{(k,2)}(e) \geq D_\beta^{(k,1)}(e) \geq D_\beta^{(k)}(e), \quad \forall e \in \mathbb{Z}.$$

Now, suppose the result of Lemma C.1 is true for some $(m-1)$, where $0 < m < m^\circ$. For any $e \in A^{(m)}$

$$D_\beta^{(k,m+1)}(e) = (1-\beta) d(e) + \beta \sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k,m)}(n), \tag{61}$$

$$D_\beta^{(k)}(e) = \begin{cases} (1-\beta) d(e) + \beta \sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k)}(n), \\ \qquad\qquad\qquad\qquad e \in A^{(m)} \setminus A^{(0)} \\ \beta \sum_{n=-\infty}^{\infty} p_n D_\beta^{(k)}(n), \qquad e \in A^{(0)}. \end{cases} \tag{62}$$

Consider any $e \in A_+^{(m)}$. If $e \in A_+^{(m-1)}$, then by monotonic increasing property of $\mathcal{T}^{(k+1)}$,

$$D_\beta^{(k,m+1)}(e) \geq D_\beta^{(k,m)}(e) > D_\beta^{(k)}(e), \tag{63}$$

where the last inequality follows from the induction hypothesis. If $e \notin A_+^{(m-1)}$, then

- $(e+b) \in A_+^{(m-1)} \implies D_\beta^{(k,m)}(e+b) > D_\beta^{(k)}(e+b)$,
- Note that $p_b > 0$, which implies that $p_b D_\beta^{(k,m)}(e+b) > p_b D_\beta^{(k)}(e+b)$.

Therefore,

$$\sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k,m)}(n) > \sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k)}(n). \tag{64}$$

Combining (61), (62), (63) and (64), we get

$$D_\beta^{(k,m+1)}(e) > D_\beta^{(k)}(e), \quad \forall e \in A_+^{(m)} \setminus A^{(0)}. \tag{65}$$

Furthermore, by (56) and monotonic increasing property of $\mathcal{T}^{(k+1)}$, we have

$$D_\beta^{(k,m+1)}(e) \geq D_\beta^{(k,1)}(e) > D_\beta^{(k)}(e), \quad \forall e \in A^{(0)}. \tag{66}$$

Combining (65) and (66), we get that

$$D_\beta^{(k,m+1)}(e) > D_\beta^{(k)}(e), \quad \forall e \in A_+^{(m)}.$$

Using a similar argument as above, we can also show that the above inequality holds for $e \in A_-^{(m)}$. Also, by monotonic increasing property of $\mathcal{T}^{(k+1)}$, we have that $D_\beta^{(k,m+m^\circ+1)}(e) \geq D_\beta^{(k,m+m^\circ)}(e) \geq D_\beta^{(k)}(e)$, $\forall e \in \mathbb{Z}$. This completes the induction step.

Hence, by principle of induction, Lemma C.1 is true for Model A.

Appendix D Proof of Lemma C.2

We show the result for Model A. The result for Model B follows similarly. We prove the result using induction. It is easy to see that by Lemma C.1, the statements of the lemma are true for $m = 0$. For $m = 1$, note that $B^{(m-1)} = \phi$ and hence $B^{(m-1)} \cup A^{(m^\circ)} = A^{(m^\circ)}$. By monotonic increasing property of $\mathcal{T}^{(k+1)}$ and Lemma C.1, we have the following:

$$D_\beta^{(k,m^\circ+2)} \geq D_\beta^{(k,m^\circ+1)} > D_\beta^{(k)}, \quad \forall e \in A^{(m^\circ)},$$

which is the result of the lemma for $m = 1$. Now, let us assume that Lemma C.2 is true for some integer $m > 1$, i.e.

$$D_\beta^{(k,m+m^\circ)}(e) > D_\beta^{(k)}(e), \quad \forall e \in B^{(m-1)} \cup A^{(m^\circ)}. \tag{67}$$

Now, consider $e \in B_+^{(m)}$. If $e \in B_+^{(m-1)}$, then by monotonic increasing property of $\mathcal{T}^{(k+1)}$,

$$D_\beta^{(k,m+m^\circ+1)}(e) \geq D_\beta^{(k,m+m^\circ)}(e) > D_\beta^{(k)}(e),$$

where the last inequality follows from the induction hypothesis. If $e \notin B_+^{(m-1)}$, then

- $(e-b) \in B_+^{(m-1)} \implies D_\beta^{(k,m+m^\circ)}(e-b) > D_\beta^{(k)}(e-b)$,
- Note that $p_b > 0$, which implies that $p_b D_\beta^{(k,m)}(e-b) > p_b D_\beta^{(k)}(e-b)$.

Thus,

$$\sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k,m+m^\circ)}(n) > \sum_{n=-\infty}^{\infty} p_{n-e} D_\beta^{(k)}(n). \tag{68}$$

Combining (52), (53) and (68) we get

$$D_\beta^{(k,m+m^\circ+1)}(e) > D_\beta^{(k)}(e), \quad \forall e \in B_+^{(m)} \cup A^{(m^\circ)}.$$

Proceeding in a similar way as above, it can be shown that the above inequality holds for all $e \in B_-^{(m)} \cup A^{(m^\circ)}$. Also, by monotonic increasing property of $\mathcal{T}^{(k+1)}$, we have that $D_\beta^{(k,m+m^\circ+1)}(e) \geq D_\beta^{(k,m+m^\circ)}(e) \geq D_\beta^{(k)}(e)$, $\forall e \in \mathbb{Z}$. This completes the induction step. Hence, by principle of induction, Lemma C.2 is true.