OPTIMAL SENSOR SCHEDULING AND REMOTE ESTIMATION OVER AN ADDITIVE NOISE CHANNEL

BY

XIAOBIN GAO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Adviser:

Professor Tamer Başar

ABSTRACT

In the applications of wireless sensor networks, sensors are built to measure the state of the system of interest and send their measurements to a remote decision unit via wireless communication. Based on the messages received from the sensors, the decision unit estimates the state of the system and makes decisions. In this scenario, the quality of decision making strongly depends on the quality of state estimation. On the other hand, the sensors are constrained by limited power and cannot always communicate with the decision unit. As a consequence, a communication scheduling strategy and an estimation strategy should be designed for the sensors and the decision unit, respectively, such that the state estimation error is minimized under the communication constraints. In this thesis, we consider a sensor scheduling and remote estimation problem with one sensor and one estimator. The sensor makes a series of observations on the state of a source and then decides whether to transmit each one in the sequence to the estimator. The sensor is charged a cost for each transmission. The remote estimator generates real-time estimates on the state of the source based on the messages received from the sensor. The estimator is charged for estimation error. In contrast to prior work in the literature, we further assume that there is additive communication channel noise, which makes the problem more challenging. As a consequence of the presence of channel noise, the sensor needs to encode the message before transmitting it to the estimator. For some specific distributions of the underlying random variables, we obtain a person-by-person optimal solution to the problem of minimizing the expected value of the sum of communication cost and estimation cost over the time horizon, which is globally optimal in the asymptotic case. In a modified problem we show that our solution is locally optimal and a globally optimal solution exists.

To my parents, for their love and support.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The sensor scheduling and remote state estimation problem arises in the applications of wireless sensor networks, such as environmental monitoring and networked control systems. As an example of environmental monitoring, people in the National Aeronautics and Space Administration (NASA) Earth Science group want to monitor the evolution of the soil moisture, which is used in weather forecasting, ecosystem process simulation and so on [1]. In order to achieve that goal, the sensor networks are built over an area of interest. The sensors collect data on the soil moisture and send them to the decision unit at NASA via wireless communication. The decision unit at NASA forms estimates of the evolution of the soil moisture based on the messages received from the sensors. Similarly, in networked control systems, the objective is to control some remote plants. Sensor networks are built to measure the states of the remote plants and then transmit their measurements to the controller via a wireless communication network. The controller estimates the state of the remote plant and generates the control signal based on that estimate [2]. In both scenarios, the quality of the remote state estimation strongly affects the quality of decision making at the remote site, that is, weather prediction or control signal generation. The networked sensors are usually constrained by limits on power [3, 4]. They are not able to communicate with the estimator at every time step and thus, the estimator has to produce its best estimate based on the partial information received from the sensors. Therefore, the communication between the sensors and the estimator should be scheduled wisely, and the estimator should be designed properly, so that the state estimation error is minimized subject to the communication constraints.

1.1 Literature Review

Research on the general sensor scheduling problem dates back to the 1970s. In one of the earliest works [5], the problem formulation was such that only one out of several sensors could be selected at each instant of time to observe the output of a linear stochastic system. After taking the observations over a finite time interval, one needs to predict some future state of the system. Furthermore, each sensor was associated with a certain measurement cost. The author proposed an off-line deterministic sensor scheduling strategy that minimizes the sum of the measurement cost over the time interval and the prediction error. Gupta *et al.* [6] studied the sensor scheduling problem over infinite time horizon. Similar to the problem in [5], only one sensor could be selected at each instant of time. However, in their formulation there is no measurement cost associated with each sensor. The authors of [6] proposed an off-line stochastic sensor scheduling strategy such that the expected steady state estimation error is minimized. Yang and Shi [7] studied the off-line sensor scheduling problem where there is only one sensor observing the state of a linear stochastic system. The sensor was able to communicate with the remote estimator for only a limited number of times. The objective was to minimize the cumulative estimation error over a finite time horizon. It was shown that the optimal sensor scheduling strategy is to distribute the limited communication opportunities uniformly over the time horizon.

The authors of the papers discussed above considered off-line sensor scheduling problems. "Off-line sensor scheduling" means the sensor is scheduled to take an observation or conduct communication based on some priori information about the system (e.g. statistics of random variables, system matrices). The on-line information (e.g. sensor's observation, battery's energy level) is not taken into account when making schedules. More work on off-line sensor scheduling problems can be found in [8, 9, 10].

Hardware development has endowed sensors with stronger computational capability. As a consequence, the sensors are able to make schedules based on all the information they have (a priori information and on-line information), which motivates people to study on-line sensor scheduling problems. Åström and Bernhardsson [11] considered a state estimation problem with a first-order stochastic system. They compared the estimation error over infinite-time horizon obtained by periodic sampling and threshold event-triggered

sampling. The periodic sampling is one of the off-line sensor scheduling strategies while the threshold event-triggered sampling is one of the on-line sensor scheduling strategies. They showed that the threshold event-triggered sampling, which is also called "threshold-based communication strategy," gives better performance in state estimation. The global optimality of a threshold-based communication strategy has been proved later by Nar and Başar [12]. Imer and Başar [13, 14] considered the on-line sensor scheduling and remote state estimation problem over a finite-time horizon. The sensor is allowed to communicate only a limited number of times. By restricting the communication strategies to the class of threshold-based ones, the authors showed that there exists a unique threshold-based communication strategy that achieves the best performance in remote state estimation. Furthermore, the optimal threshold can be computed by solving a dynamic programming equation, which is discussed in [15]. Bommannavar and Başar [16] later extended the result of [13] to a multi-dimensional systems. The continuous-time version of the problem in [13] has been studied by Rabi *et al.* [17, 18]. Xu and Hespanha [19] considered the networked control problem involving state estimation and communication scheduling, which can be viewed as a sensor scheduling and remote estimation problem. They fixed the estimator to be a Kalman-like estimator and designed an event-triggered sensor that minimizes the time average of the sum of the communication cost and estimation error over infinite horizon. They showed that the optimal communication strategy is deterministic and stationary, and is a function of the estimation error. Wu *et al.* [20] considered the sensor scheduling and estimation problem subject to constraints on the average communication rate over the infinite horizon. The authors assumed that the sensor has noisy observations on the system state. By restricting the sensor scheduling strategies to the threshold event-triggered class, they derived the exact minimum mean square error (MMSE) estimator. However, the exact MMSE estimator is nonlinear and thus computationally intractable. By making the Gaussian assumption on the a priori distribution, the authors derived an approximate MMSE estimator, which is a Kalman-like estimator. Based on the approximated MMSE estimator, the authors derived the conditions on the thresholds so that the average sensor communication rate will not exceed its upper bound. Note that making a Gaussian assumption to derive an approximate MMSE estimator is a widely used technique when studying nonlinear filtering problems [21, 22, 23]. You

and Xie [24] extended the work in [20] by deriving the conditions on the thresholds so that the estimator is stable. A recent work by Han *et al.* [25] showed that if the sensor is fixed to apply some stochastic event-triggered strategy, then the exact MMSE estimator is a Kalman-like estimator. The work in [20, 24, 25] can also be viewed as Kalman-filtering with scheduled observations, which is related to Kalman-filtering with intermittent observations studied in [26, 27].

The approaches of [19, 20] were to fix the communication strategies or estimation strategies to be of a certain type and to derive the corresponding optimal estimation strategies and communication strategies, respectively. The approach of [13], on the other hand, is to derive the optimal communication strategies and estimation strategies simultaneously. Similarly, Lipsa and Martins [28] considered the sensor scheduling and remote estimation problem where the sensor is not constrained by communication times but is charged a communication cost. They proposed a threshold event-triggered sensor and a Kalman-like estimator and proved that the proposed sensor and estimator are jointly optimal, minimizing the sum of communication cost and estimation error over a finite-time horizon. Nayyar *et al.* [29] considered a similar problem where the sensor is equipped with an energy harvesting sensor. In the work of [29], the problem formulation is such that the sensor is constrained by the energy level of the battery and is also charged a communication cost. It is shown in [29] that an energy dependent threshold event-triggered sensor and a Kalman-like estimator are jointly optimal. Hence, the result of [29] can be viewed as a generalization of the results of [13, 28]. In both works of [28] and [29], the authors applied majorization theory to prove the optimality of their results, which is closely related to the work in [30].

Note that the problem setups in [13] and [28] are quite similar except for the communication constraint. In the problem of [13], the sensor can only communicate for a pre-specified number of times. Such a communication constraint is called *hard constraint*. In the work of [28], however, the sensor is charged a communication cost. This kind of communication constraint is called *soft constraint*. In the problem with hard constraint, the communication strategy must take the remaining communication opportunities as a variable and schedule no communication if there is no remaining opportunity. Such communication strategies guarantee that the number of transmissions made over the time horizon will not exceed the given constraint. In the

problem with soft constraint, however, the sensor is not constrained by the number of transmissions. Therefore, the communication strategy need not take the remaining communication opportunities (which are always equal to the remaining time steps) as a variable. Therefore, the results obtained in one problem cannot be applied directly to the other. For example, if we apply the communication strategy obtained in [28] to the problem of [13], then there exists a positive probability that the sensor decides to communicate at every instant of time, which certainly violates the hard constraint on communication times. A detailed discussion of the difference between optimization problems with soft constraints and those with hard constraints can be found in [31, 32].

1.2 Contributions

In prior work, communication between the sensor and the estimator has been assumed to be perfect (no additive channel noise), which may not be the case in reality. Hence we are motivated to investigate the effect of communication channel noise on the design of optimal sensor scheduling and remote estimation strategies. Accordingly, we consider in this thesis a discrete-time sensor scheduling and remote estimation problem over a finite-time horizon, where there is one sensor and one remote estimator. We assume that at each time step, the sensor makes a perfect observation of the state of an independent identically distributed (i.i.d.) source. Then the sensor decides whether to transmit its observation to the remote estimator or not. The sensor is charged a cost for each transmission (communication cost). Since the communication channel is noisy, the sensor encodes the message before transmitting it to the estimator. The remote estimator generates a real-time estimate of the state of the source based on the noise-corrupted messages received from the sensor. The estimator is charged for estimation error (estimation cost). Our goal is to design the communication scheduling strategy and encoding strategy for the sensor, and the estimation strategy (decoding strategy) for the estimator, to minimize the expected value of the sum of communication cost and estimation cost over the time horizon. We propose a solution to this problem, which consists of a threshold-based communication scheduling strategy, and a pair of piecewise linear encoding/decoding

strategies. We show that the proposed solution is person-by-person optimal when the random variables have some specific distributions. We also show that our solution is globally optimal in the asymptotic case. In a modified problem we show the local optimality of our solution and the existence of a globally optimal solution. Note that the optimization problem without sensor scheduling and communication cost has been discussed in [33, 34, 35].

1.3 Organizations

The rest of the thesis is organized as follows: In chapter 2 we describe the system model, formulate the optimization problem and state the assumptions on parameters. In chapter 3 we review and extend some results from [35], which establish the necessary and sufficient conditions on the distributions of source and channel noise such that the optimal encoder and decoder are linear. The necessary and sufficient conditions are used in the proof of the main results. In chapter 4 we present and prove the main results. Finally in chapter 5 we make conclusions and discuss future directions.

CHAPTER 2

PROBLEM FORMULATION

2.1 System Model

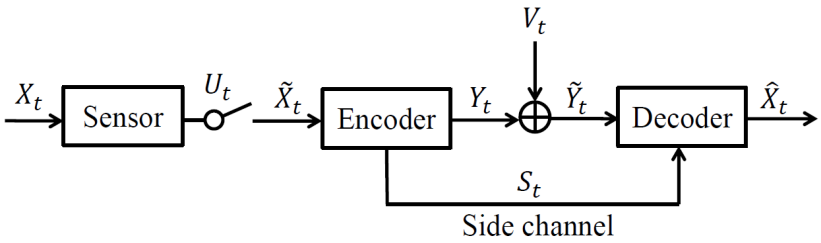

Figure 2.1: System model

Consider a discrete time sensor scheduling and remote estimation problem over a finite time horizon, that is, $t = 1, 2, \ldots, T$. In the problem, there is *one* sensor, *one* encoder and *one* remote estimator (which is also called "decoder"), as illustrated in Fig. 2.1. The sensor observes an independent identically distributed (i.i.d.) stochastic process, $\{X_t\}$, $X_t \in \mathbb{R}$, which has Laplace distribution with parameters $(0, \lambda^{-1})$. Hence,

$$f_X(x) = \begin{cases} \frac{1}{2}\lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ \frac{1}{2}\lambda e^{\lambda x}, & \text{if } x < 0 \end{cases}$$

Assume that at time t the sensor has perfect observation on X_t. Then the sensor decides whether to transmit the measurement to the encoder or not. Let $U_t \in \{0, 1\}$ be the sensor's decision at time t, where $U_t = 0$ stands for no communication and $U_t = 1$ stands for communication. If the sensor communicates at time t, it will be charged a cost c. "No communication" results in zero communication cost. Assume that the communication between the sensor and the encoder is perfect. Denote by $\tilde{X}_t$ the message received by

the encoder. Then,

$$\tilde{X}_t = \begin{cases} X_t, & \text{if } U_t = 1 \\ \epsilon, & \text{if } U_t = 0 \end{cases}$$

where ϵ is a free symbol standing for no message is transmitted. Once the encoder receives the message from the sensor, it sends an encoded message to the decoder, denoted by Y_t. The encoder will not send any message to the decoder if it does not receive any message from the sensor, which is denoted by $Y_t = \epsilon$. Assume that the encoded message is corrupted by an additive channel noise V_t, $V_t \in \mathbb{R}$. $\{V_t\}$ is an i.i.d. random process, which is independent of $\{X_t\}$. We take V_t to have gamma distribution $\Gamma(k, \theta)$. Denoting the message received by the decoder by $\tilde{Y}_t$, we have

$$\tilde{Y}_t = \begin{cases} Y_t + V_t, & \text{if } Y_t \neq \epsilon \\ \epsilon, & \text{if } Y_t = \epsilon \end{cases}$$

When sending the encoded message to the decoder, the encoder will transmit the sign of $\tilde{X}_t$ to the decoder via a side channel, denoted by S_t. Again, the encoder will not send any message to the decoder via the side channel if it does not receive any message from the sensor. Assume that the side channel is noise-free. Then,

$$S_t = \begin{cases} \text{sgn}(\tilde{X}_t), & \text{if } \tilde{X}_t \neq \epsilon \\ \epsilon, & \text{if } \tilde{X}_t = \epsilon \end{cases}$$

After receiving $\tilde{Y}_t$ and S_t, the decoder produces an estimate on X_t, denoted by $\hat{X}_t$. The estimator will be charged for distortion in estimation. Assume that the distortion function $\rho(X_t, \hat{X}_t)$ is the squared error $(X_t - \hat{X}_t)^2$.

Assume that the encoder has average power constraint, that is

$$\mathbb{E}\left[Y_t^2 \,\middle|\, U_t = 1\right] \leq P_T$$

where P_T is known. The cost incurred by the system at time t is

$$cU_t + \left(X_t - \hat{X}_t\right)^2, \quad c > 0$$

where cU_t is the communication cost and $(X_t - \hat{X}_t)^2$ is the estimation cost.

2.2 Decision Strategies

Assume that at time t, the sensor has memory on all its observations up to t, denoted by $X_{1:t}$, and all the decisions it has made up to $t-1$, denoted by $U_{1:t-1}$. The sensor determines whether to communicate or not at time t based on its current information $(X_{1:t}, U_{1:t-1})$, namely

$$U_t = f_t(X_{1:t}, U_{1:t-1})$$

where f_t is the scheduling policy of the sensor at time t and $\mathbf{f} = \{f_1, f_2, \ldots, f_T\}$ is the scheduling strategy of the sensor.

Similarly, at time t, the encoder has memory on all the messages received from the sensor up to t, denoted by $\tilde{X}_{1:t}$, and all the encoded messages it has sent to the decoder up to $t-1$, denoted by $Y_{1:t-1}$. The encoder generates the encoded message at time t based on its current information $(\tilde{X}_{1:t}, Y_{1:t-1})$, namely

$$Y_t = g_t(\tilde{X}_{1:t}, Y_{1:t-1})$$

where g_t is the encoding policy of the encoder at time t and $\mathbf{g} = \{g_1, g_2, \ldots, g_T\}$ is the encoding strategy.

Finally, assume that at time t, the decoder has memory of all the messages received from the encoder up to t, denoted by $\tilde{Y}_{1:t}, S_{1:t}$. The decoder generates the estimate at time t based on its current information $(\tilde{Y}_{1:t}, S_{1:t})$, namely

$$\hat{X}_t = h_t(\tilde{Y}_{1:t}, S_{1:t})$$

where h_t is the decoding policy of the decoder at time t and $\mathbf{h} = \{h_1, h_2, \ldots, h_T\}$ is the decoding strategy.

Remark 1. *Although we do not assume that the encoder has memory on $U_{1:t}, S_{1:t}$, yet it can deduce them from $\tilde{X}_{1:t}$. Similarly, the decoder can obtain $U_{1:t}$ from $\tilde{Y}_{1:t}$.*

2.3 Assumptions on Parameters

Denote by σ_V^2 the variance of V_t, and recall that V_t has gamma distribution $\Gamma(k, \theta)$. Then, $\sigma_V^2 = k\theta^2$. Define $\alpha =: \lambda\sqrt{P_T}$, and $\gamma := \frac{P_T}{\sigma_V^2}$. Assume that

$$\theta = \sqrt{P_T}$$

Then we have

$$\begin{aligned} \alpha &= \lambda\theta \\ \gamma &= \frac{1}{k} \end{aligned} \tag{2.1}$$

2.4 Optimization Problem

Consider the system described above, given the time horizon T, the statistics of $\{X_t\}$ and $\{V_t\}$, the communication cost c, and the power constraint P_T. Determine the scheduling strategy, encoding strategy and decoding strategy $(\mathbf{f}, \mathbf{g}, \mathbf{h})$ for the sensor, the encoder and the decoder, respectively, that minimize the expected value of the sum of communication cost and estimation cost over the time horizon, namely,

$$J(\mathbf{f}, \mathbf{g}, \mathbf{h}) = \mathbb{E}\left\{\sum_{t=1}^{T} cU_t + \left(X_t - \hat{X}_t\right)^2\right\}$$

subject to the power constraint of the encoder.

CHAPTER 3

PRIOR WORK

3.1 Existing Results on Zero Delay Source-Channel Coding

We first review some results of a communication problem studied by Akyol *et al* [35]. Consider the communication system described in Fig. 3.1. The

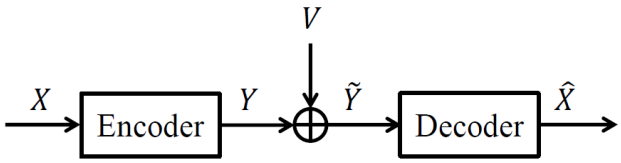

Figure 3.1: Communication system considered in prior work

encoder wants to transmit an input signal X to the decoder via a communication channel. X is a random variable, $X \in \mathbb{R}$. The communication channel has an additive channel noise V. V is also a random variable taking values in $\mathbb{R}$. Assume that X and V are independent, whose characteristic functions are $F_X(\omega)$ and $F_V(\omega)$, respectively. Denote by σ_X^2 and σ_V^2 the variances of X and V, respectively. Since the communication channel is noisy, the encoder needs to encode the message before sending it to the decoder. Assume that the encoder generates the encoded message Y according to some encoding policy g, namely

$$Y = g(X)$$

The encoder is constrained by average power P_T such that

$$\mathbb{E}[Y^2] \leq P_T$$

The decoder receives the noise corrupted message $Y + V$, denoted by $\tilde{Y}$. Then the decoder generates an estimate of X, denoted by $\hat{X}$, according to

some decoding policy h, and hence

$$\hat{X} = h(\tilde{Y})$$

Communication problem: Given the distributions of X and V, and the power constraint of the encoder P_T, design the encoding and decoding policies g and h, respectively, that minimize the mean squared estimation error,

$$J(g,h) = \mathbb{E}\left[(X - \hat{X})^2\right]$$

subject to the power constraint of the encoder.

According to Akyol *et al.* [35], the communication problem above has the following solution.

Theorem 1. *Consider the communication problem described above, and assume that X and V are zero mean random variables.*

1. *The optimal encoder and decoder are either both linear or both nonlinear.*

2. *The optimal encoder and decoder are both linear if and only if the characteristic functions of X and V satisfy*

$$F_X(\alpha\omega) = F_V^{\gamma}(\omega)$$

where $\alpha = \sqrt{\frac{P_T}{\sigma_X^2}}$ and $\gamma = \frac{P_T}{\sigma_V^2}$. Moreover, the linear encoding/decoding policies (g^*, h^*) are as follows:

$$\begin{aligned} Y &= g^*(X) = \alpha X \\ \hat{X} &= h^*(\tilde{Y}) = \frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y} \end{aligned} \tag{3.1}$$

3.2 Extended Results on Zero Delay Source-Channel Coding

We apply Theorem 1 to a special case relevant to the problem of this thesis, and arrive at the following lemma.

Lemma 1. *Consider the communication problem described above, assume that X and V can be written as*

$$\begin{aligned} X &= X_e - \lambda^{-1} \\ V &= V_g - k\theta \end{aligned}$$

where X_e has exponential distribution with parameter λ, and V_g has gamma distribution with parameters (k, θ). Furthermore, let θ and P_T satisfy $\theta = \sqrt{P_T}$. Then the optimal encoding/decoding policies (g^*, h^*) are as described in (3.1) with $\alpha = \lambda\sqrt{P_T}$ and $\gamma = \frac{P_T}{k\theta^2}$.

Proof. From the definitions of X and V, we have

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_e] - \lambda^{-1} = 0 \\ \mathbb{E}[V] &= \mathbb{E}[V_g] - k\theta = 0 \\ \sigma_X^2 &= \mathrm{Var}(X_e) = \lambda^{-2} \\ \sigma_V^2 &= \mathrm{Var}(V_g) = k\theta^2 \end{aligned}$$

Then

$$\begin{aligned} \alpha &= \sqrt{\tfrac{P_T}{\sigma_X^2}} = \lambda\sqrt{P_T} \\ \gamma &= \tfrac{P_T}{\sigma_V^2} = \tfrac{P_T}{k\theta^2} \end{aligned}$$

By the assumption on the parameters (2.1), we have

$$\begin{aligned} F_X(\alpha\omega) &= \mathbb{E}\left[e^{j\alpha\omega X}\right] \\ &= \mathbb{E}\left[e^{j\alpha\omega(X_e - \lambda^{-1})}\right] \\ &= \mathbb{E}\left[e^{j\alpha\omega X_e}\right] e^{-j\alpha\omega\lambda^{-1}} \\ &= (1 - j\alpha\omega\lambda^{-1})^{-1} e^{-j\alpha\omega\lambda^{-1}} \\ &= (1 - j\omega\theta)^{-1} e^{-j\omega\theta} \end{aligned}$$

Similarly,

$$\begin{aligned} F_V(\omega) &= \mathbb{E}\left[e^{j\omega V}\right] \\ &= \mathbb{E}\left[e^{j\omega(V_g - k\theta)}\right] \\ &= \mathbb{E}\left[e^{j\omega V_g}\right] e^{-j\omega k\theta} \\ &= (1 - j\omega\theta)^{-k} e^{-j\omega k\theta} \\ &= \left[(1 - j\omega\theta)^{-1} e^{-j\omega\theta}\right]^k \end{aligned}$$

Hence,

$$\begin{aligned} F_V^{\gamma}(\omega) &= [F_V(\omega)]^{\frac{1}{k}} \\ &= (1 - j\omega\theta)^{-1} e^{-j\omega\theta} \\ &= F_X(\alpha\omega) \end{aligned}$$

Applying Theorem 1, the optimal encoding/decoding policies g^*, h^* are linear, and are described by (3.1).

□

We now prove a lemma such that the result of Theorem 1 can be extended to the case where X and V are not zero mean random variables but a linear transform of zero mean random variables.

Lemma 2. *Consider the communication problem described above with zero mean random variables X and V; let us call it Problem 1 . Suppose that given the distributions of X and V, and the power constraint of the encoder P_T, the optimal encoding/decoding policies to Problem 1 are (g^*, h^*). Then consider the communication problem described above with the same power constraint but non-zero mean random variables X' and V'; call it Problem 2. Assume that X' and V' are linear transforms of X and V, namely*

$$\begin{aligned} X' &= aX + b_1 \\ V' &= V + b_2 \end{aligned}$$

where a, b_1, b_2 are known constants, $a \in \{-1, +1\}$, $b_1, b_2 \in \mathbb{R}$. Then the optimal encoding/decoding policies to Problem 2, denoted by (g'^*, h'^*), can be obtained by

$$\begin{aligned} g'^*(X') &= g^*\left(\tfrac{X' - b_1}{a}\right) \\ h'^*(\tilde{Y}') &= a \cdot h^*(\tilde{Y}' - b_2) + b_1 \end{aligned}$$

Moreover, the optimal costs of the two problems are equivalent.

Proof. For any encoding/decoding policies (g, h) satisfying the power constraint in *Problem 1*, we have

$$\begin{aligned} Y &= g(X) \\ \mathbb{E}[Y^2] &\leq P_T \\ \hat{X} &= h(\tilde{Y}) = h(Y + V) \end{aligned}$$

Define the encoding/decoding policies (g', h') in *Problem 2* as follows:

$$\begin{aligned} Y' = g'(X') &= g\left(\tfrac{X'-b_1}{a}\right) \\ \hat{X}' = h'(\tilde{Y}') &= a \cdot h(\tilde{Y}' - b_2) + b_1 \end{aligned}$$

where Y' is the output of the encoder, $\tilde{Y}'$ is the noise corrupted message received by the decoder, and $\hat{X}'$ is the output of the decoder in *Problem 2*. Furthermore,

$$\tilde{Y}' = Y' + V'$$

Then we have

$$\begin{aligned} Y' &= g'(X') \\ &= g\left(\tfrac{X'-b_1}{a}\right) \\ &= g(X) \\ &= Y \end{aligned}$$

Hence,

$$\mathbb{E}[Y'^2] = \mathbb{E}[Y^2] \leq P_T$$

which implies that the pair (g', h') satisfies the power constraint of the encoder in *Problem 2*. Moreover,

$$\begin{aligned} \hat{X}' &= h'(\tilde{Y}') \\ &= a \cdot h(\tilde{Y}' - b_2) + b_1 \\ &= a \cdot h(Y' + V' - b_2) + b_1 \\ &= a \cdot h(Y + V) + b_1 \\ &= a\hat{X} + b_1 \end{aligned}$$

Let $J_1(g, h)$ be the cost corresponding to (g, h) in *Problem 1* , and $J_2(g', h')$

be the cost corresponding to (g', h') in *Problem 2*. Then,

$$\begin{aligned} J_2(g', h') &= \mathbb{E}\left[(X' - \hat{X}')^2\right] \\ &= \mathbb{E}\left[(aX + b_1 - a\hat{X} - b_1)^2\right] \\ &= a^2 \cdot \mathbb{E}\left[(X - \hat{X})^2\right] \\ &= \mathbb{E}\left[(X - \hat{X})^2\right] \\ &= J_1(g, h) \end{aligned}$$

Conversely, for any pair of encoding/decoding policies (g', h') satisfying the power constraint in *Problem 2*, define the encoding/decoding policies (g, h) in *Problem 1* as follows:

$$\begin{aligned} Y = g(X) &= g'\left(aX + b_1\right) \\ \hat{X} = h(\tilde{Y}) &= \frac{1}{a}\left[h'(\tilde{Y} + b_2) - b_1\right] \end{aligned}$$

Then we have

$$\begin{aligned} Y &= g(X) \\ &= g'\left(aX + b_1\right) \\ &= g'(X') \\ &= Y' \end{aligned}$$

Hence,

$$\mathbb{E}[Y^2] = \mathbb{E}[Y'^2] \leq P_T$$

which implies that the pair (g, h) satisfies the power constraint of the encoder in *Problem 1*. Moreover,

$$\begin{aligned} \hat{X} &= h(\tilde{Y}) \\ &= \frac{1}{a}\left[h'(\tilde{Y} + b_2) - b_1\right] \\ &= \frac{1}{a}\Big[h'(Y + V + b_2) - b_1\Big] \\ &= \frac{1}{a}\Big[h'(Y' + V') - b_1\Big] \\ &= \frac{1}{a}\left(\hat{X}' - b_1\right) \end{aligned}$$

Then,

$$\begin{aligned} J_1(g,h) &= \mathbb{E}\left[\left(X-\hat{X}\right)^2\right] \\ &= \mathbb{E}\left[\left(\frac{1}{a}(X'-b_1)-\frac{1}{a}(\hat{X}'-b_1)\right)^2\right] \\ &= \frac{1}{a^2}\cdot\mathbb{E}\left[\left(X'-\hat{X}'\right)^2\right] \\ &= \mathbb{E}\left[\left(X'-\hat{X}'\right)^2\right] \\ &= J_2(g',h') \end{aligned}$$

Therefore, the optimal costs of *Problem 1* and *Problem 2* are equivalent. Moreover, if the optimal cost of *Problem 1* is achieved by (g^*,h^*), then the optimal cost of *Problem 2* can be achieved by

$$\begin{aligned} g'^*(X') &= g^*\left(\tfrac{X'-b_1}{a}\right) \\ h'^*(\tilde{Y}') &= a\cdot h^*(\tilde{Y}'-b_2)+b_1 \end{aligned}$$

□

Applying Lemma 2 to Lemma 1, we have the following result.

Lemma 3. *Consider the communication problem described above, assume that the input signal X has exponential distribution with parameter λ, and communication channel noise V has gamma distribution with parameters (k,θ). Furthermore, let θ, P_T satisfy $\theta=\sqrt{P_T}$. Then the optimal encoding/decoding policies (g^*,h^*) are as follows:*

$$\begin{aligned} Y &= g^*(X)=\alpha X-\alpha\lambda^{-1} \\ \hat{X} &= h^*(\tilde{Y})=\frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y}+\frac{\gamma}{\gamma+1}\lambda^{-1} \end{aligned}$$

where $\alpha=\lambda\sqrt{P_T}$ and $\gamma=\frac{P_T}{k\theta^2}$. Moreover, the optimal cost is

$$J(g^*,h^*)=\frac{1}{\gamma+1}\frac{1}{\lambda^2}:=m$$

Proof. By Lemma 1, if the input signal X' and the communication channel noise V' are as follows:

$$\begin{aligned} X' &= X-\lambda^{-1} \\ V' &= V-k\theta \end{aligned}$$

Then the optimal encoding/decoding policies are

$$\begin{aligned} Y' &= g'^*(X') = \alpha X' \\ \hat{X}' &= h'^*(\tilde{Y}') = \frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y}' \end{aligned}$$

From the definitions of X' and V', we have

$$\begin{aligned} X &= X' + \lambda^{-1} \\ V &= V' + k\theta \end{aligned}$$

Applying Lemma 2 by letting $a = 1, b_1 = \lambda^{-1}, b_2 = k\theta$, we have the optimal encoding/decoding policies corresponding to X and V are as follows:

$$\begin{aligned} Y &= g^*(X) \\ &= g'^*\left(\tfrac{X-b_1}{a}\right) \\ &= \alpha(X - \lambda^{-1}) \\ &= \alpha X - \alpha\lambda^{-1} \\ \hat{X} &= h^*(\tilde{Y}) \\ &= a \cdot h'^*(\tilde{Y} - b_2) + b_1 \\ &= \frac{1}{\alpha}\frac{\gamma}{\gamma+1}(\tilde{Y} - k\theta) + \lambda^{-1} \\ &= \frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y} + \frac{\gamma}{\gamma+1}\lambda^{-1} \end{aligned}$$

Moreover,

$$\begin{aligned} \hat{X} &= \frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y} + \frac{\gamma}{\gamma+1}\lambda^{-1} \\ &= \frac{1}{\alpha}\frac{\gamma}{\gamma+1}(Y + V) + \frac{\gamma}{\gamma+1}\lambda^{-1} \\ &= \frac{1}{\alpha}\frac{\gamma}{\gamma+1}(\alpha X - \alpha\lambda^{-1} + V) + \frac{\gamma}{\gamma+1}\lambda^{-1} \\ &= \frac{\gamma}{\gamma+1}X + \frac{1}{\alpha}\frac{\gamma}{\gamma+1}V \end{aligned}$$

By the assumptions on the parameters (2.1), the cost of (g^*, h^*) is computed

as follows:

$$
\begin{aligned}
& J(g^*, h^*) \\
= \;& \mathbb{E}\left[(X - \hat{X})^2\right] \\
= \;& \mathbb{E}\left[\left(X - \frac{\gamma}{\gamma+1}X - \frac{1}{\alpha}\frac{\gamma}{\gamma+1}V\right)^2\right] \\
= \;& \mathbb{E}\left[\left(\frac{1}{\gamma+1}X - \frac{1}{\alpha}\frac{\gamma}{\gamma+1}V\right)^2\right] \\
= \;& \frac{1}{(\gamma+1)^2}\mathbb{E}\left[\left(X - \alpha^{-1}\gamma V\right)^2\right] \\
= \;& \frac{1}{(\gamma+1)^2}\Big(\mathbb{E}[X^2] + \alpha^{-2}\gamma^2\mathbb{E}[V^2] - 2\alpha^{-1}\gamma\mathbb{E}[XV]\Big) \\
= \;& \frac{1}{(\gamma+1)^2}\Big(\mathbb{E}[X]^2 + \mathrm{Var}(X) + \alpha^{-2}\gamma^2\mathbb{E}[V]^2 + \alpha^{-2}\gamma^2\mathrm{Var}(V) \\
& \qquad\qquad -2\alpha^{-1}\gamma\mathbb{E}[X]\mathbb{E}[V]\Big) \\
= \;& \frac{1}{(\gamma+1)^2}\Big(\lambda^{-2} + \lambda^{-2} + \alpha^{-2}\gamma^2k^2\theta^2 + \alpha^{-2}\gamma^2k\theta^2 - 2\alpha^{-1}\gamma\lambda^{-1}k\theta\Big) \\
= \;& \frac{1}{(\gamma+1)^2}\Big(2\lambda^{-2} + \lambda^{-2} + \lambda^{-2}\gamma - 2\lambda^{-2}\Big) \\
= \;& \frac{1}{\gamma+1}\frac{1}{\lambda^2} \\
:= \;& m
\end{aligned}
$$

□

CHAPTER 4

MAIN RESULTS

4.1 Stationary Property of the Optimal Solution

Theorem 2. *Consider the sensor scheduling and remote estimation problem described in chapter 2,*

1. *Without loss of optimality, one can restrict the sensor scheduling, encoding and decoding strategies to the form*

$$\begin{aligned} U_t &= f_t(X_t) \\ Y_t &= g_t(\tilde{X}_t) \\ \hat{X}_t &= h_t(\tilde{Y}_t, S_t) \end{aligned}$$

2. *There exist stationary strategies* $(\boldsymbol{f}^*, \boldsymbol{g}^*, \boldsymbol{h}^*)$ *minimizing*

$$J(\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h}) = \mathbb{E}\left\{\sum_{t=1}^{T} cU_t + (X_t - \hat{X}_t)^2\right\}$$

where

$$\begin{aligned} \boldsymbol{f}^* &= \{f_1^*, f_2^*, \ldots, f_T^*\}, \quad f_1^* = \cdots = f_T^* := f^* \\ \boldsymbol{g}^* &= \{g_1^*, g_2^*, \ldots, g_T^*\}, \quad g_1^* = \cdots = g_T^* := g^* \\ \boldsymbol{h}^* &= \{h_1^*, h_2^*, \ldots, h_T^*\}, \quad h_1^* = \cdots = h_T^* := h^* \end{aligned}$$

Moreover, (f^*, g^*, h^*) *can be obtained by solving the "one-stage" problem, that is the problem with* $T = 1$*.*

Proof. At time $t = T$, we want to design (f_T, g_T, h_T) to minimize

$$J_{T_1}(f_T, g_T, h_T) = \mathbb{E}\left\{cU_T + (X_T - \hat{X}_T)^2\right\}$$

subject to the power constraint of the encoder, called *Problem T1.* Let

I_{sT}, I_{eT}, I_{dT} be the information about the past system states available to the sensor, the encoder and the decoder, respectively, at time T. Namely

$$\begin{aligned} I_{sT} &= \{X_{1:T-1}, U_{1:T-1}\} \\ I_{eT} &= \{\tilde{X}_{1:T-1}, Y_{1:T-1}\} \\ I_{dT} &= \{\tilde{Y}_{1:T-1}, S_{1:T-1}\} \end{aligned}$$

Recall that the decisions of the sensor, the encoder and the decoder at time T are generated by

$$\begin{aligned} U_T &= f_T(X_T, I_{sT}) \\ Y_T &= g_T(\tilde{X}_T, I_{eT}) \\ \hat{X}_T &= h_T(\tilde{Y}_T, S_T, I_{dT}) \end{aligned}$$

We will show that using information about the past (I_{sT}, I_{eT}, I_{dT}) when making decisions cannot help improve the performance (reducing the expected cost).

Let I_T be the information set about the past system states at time T, namely

$$I_T = \{X_{1:T-1}, U_{1:T-1}, \tilde{X}_{1:T-1}, Y_{1:T-1}, \tilde{Y}_{1:T-1}, S_{1:T-1}\}$$

Then $I_{sT}, I_{eT}, I_{dT} \in I_T$. Consider another problem, called *Problem T2*, where I_T is available to the sensor, the encoder and the decoder and we want to design (f'_T, g'_T, h'_T) to minimize

$$J_{T_2}(f'_T, g'_T, h'_T) = \mathbb{E}\left\{cU_T + (X_T - \hat{X}_T)^2\right\}$$

subject to the power constraint of the encoder, where

$$\begin{aligned} U_T &= f'_T(X_T, I_T) \\ Y_T &= g'_T(\tilde{X}_T, I_T) \\ \hat{X}_T &= h'_T(\tilde{Y}_T, S_T, I_T) \end{aligned}$$

Since the sensor, the encoder and the decoder can always ignore the redundant information and behave as if they only know I_{sT}, I_{eT}, I_{dT}, respectively, the performance of system in *Problem T2* is no worse than the performance

of the system in *Problem T1*, namely

$$\min_{(f'_T, g'_T, h'_T)} J_{T_2}(f'_T, g'_T, h'_T) \leq \min_{(f_T, g_T, h_T)} J_{T_1}(f_T, g_T, h_T)$$

Similarly, consider a third problem, call it *Problem T3*, where I_{sT}, I_{eT}, I_{dT} are not available to the sensor, the encoder and the decoder, respectively. We want to design (f''_T, g''_T, h''_T) to minimize

$$J_{T_3}(f''_T, g''_T, h''_T) = \mathbb{E}\left\{cU_T + (X_T - \hat{X}_T)^2\right\}$$

subject to the power constraint of the encoder, where

$$\begin{aligned} U_T &= f''_T(X_T) \\ Y_T &= g''_T(\tilde{X}_T) \\ \hat{X}_T &= h''_T(\tilde{Y}_T, S_T) \end{aligned}$$

By a similar argument as above, the system in *Problem T1* cannot perform worse than the system in *Problem T3*, namely,

$$\min_{(f_T, g_T, h_T)} J_{T_1}(f_T, g_T, h_T) \leq \min_{(f''_T, g''_T, h''_T)} J_{T_3}(f''_T, g''_T, h''_T)$$

Let us come back to *Problem T2*. Since the communication cost c, the distortion function $\rho(\cdot, \cdot)$, and the power constraint of the encoder do not depend on I_T, the problem can be reformulated as a one-stage problem where the source and the communication channel noise are conditional random variables X'_T, V'_T, respectively, such that

$$\begin{aligned} X'_T &= X_T | I_T \\ V'_T &= V_T | I_T \end{aligned}$$

Since $\{X_t\}$ and $\{V_t\}$ are i.i.d. random processes, X_T and V_T are independent of I_T. Then

$$\begin{aligned} X'_T &= X_T \\ V'_T &= V_T \end{aligned}$$

Therefore, there is no loss of optimality by restricting

$$\begin{aligned} U_T &= f'_T(X_T) \\ Y_T &= g'_T(\tilde{X}_T) \\ \hat{X}_T &= h'_T(\tilde{Y}_T, S_T) \end{aligned}$$

and

$$\min_{(f'_T, g'_T, h'_T)} J_{T_2}(f'_T, g'_T, h'_T) = \min_{(f''_T, g''_T, h''_T)} J_{T_3}(f''_T, g''_T, h''_T)$$

The equality above indicates that knowing all the information about the past cannot work better than not knowing any information about the past. Hence, in *Problem T1* the sensor, the encoder and the decoder can ignore their information about the past, namely I_{sT}, I_{eT}, and I_{dT}, respectively, and there is no loss of optimality by restricting

$$\begin{aligned} U_T &= f_T(X_T) \\ Y_T &= g_T(\tilde{X}_T) \\ \hat{X}_T &= h_T(\tilde{Y}_T, S_T) \end{aligned}$$

which proves the first part of the theorem.

Since (f_T, g_T, h_T) does not take I_T as a parameter, the design of (f_T, g_T, h_T) is independent of the design of $(f_{1:T-1}, g_{1:T-1}, h_{1:T-1})$. Hence the problem can be viewed as a $T-1$ stages problem and a one stage problem.

By induction, we can show that the design of (f_1, g_1, h_1), (f_2, g_2, h_2), $\ldots$, (f_T, g_T, h_T) are mutually independent, where (f_t, g_t, h_t) is designed to minimize

$$J(f_t, g_t, h_t) = \mathbb{E}\left\{cU_t + (X_t - \hat{X}_t)^2\right\}$$

subject to the power constraint of the encoder. Furthermore, since $\{X_t\}$ and $\{V_t\}$ are i.i.d. random processes, the optimal (f_t, g_t, h_t) should be the same for all $t = 1, 2, \ldots, T$, which proves the second part of the theorem.

□

By Theorem 2, the problem can be reduced to the "one-stage" problem and the objective is to determine (f^*, g^*, h^*). Therefore for simplicity we suppress the subscript for time in all the expressions for the rest of the thesis.

4.2 Person-by-Person Optimality of the Proposed Solution

Theorem 3. *Consider the sensor scheduling and remote estimation problem formulated in chapter 2. If the sensor applies symmetric threshold-based scheduling policy f as follows:*

$$U = f(X) = \begin{cases} 1, & \text{if } |X| > \beta \\ 0, & \text{if } |X| \leq \beta \end{cases}$$

where β is called the threshold, then the encoding/decoding policies (g, h) described below are jointly optimal corresponding to f:

$$g(\tilde{X}) = \begin{cases} \alpha|\tilde{X}| - \alpha\beta - \alpha\lambda^{-1}, & \text{if } \tilde{X} \neq \epsilon \\ \epsilon, & \text{if } \tilde{X} = \epsilon \end{cases}$$

$$h(\tilde{Y}, S) = \begin{cases} S \cdot \left(\frac{1}{\alpha} \frac{\gamma}{\gamma+1} \tilde{Y} + \frac{\gamma}{\gamma+1} \lambda^{-1} + \beta \right), \text{if } \tilde{Y}, S \neq \epsilon \\ 0, & \text{if } \tilde{Y}, S = \epsilon \end{cases}$$

where $\alpha = \lambda\sqrt{P_T}$, $\gamma = \frac{P_T}{k\theta^2}$

Proof. Case I. $U = 0$, $\tilde{X}, \tilde{Y}, S = \epsilon$. The minimum mean squared error (MMSE) estimator should be the conditional mean, that is,

$$\begin{aligned} \hat{X} &= \mathbb{E}\left[X \,\middle|\, U = 0\right] \\ &= \mathbb{E}\left[X \,\middle|\, |X| \leq \beta\right] \\ &= 0 \end{aligned}$$

where the third inequality is due to the fact that X is symmetrically distributed.

Case II. $U = 1$, $S = +$. The problem collapses to the communication problem discussed in chapter 3 with the input signal $\tilde{X}$, the communication channel noise V, and the power constraint of the encoder P_T. $U = 1$, $S = +$ is equivalent to $\tilde{X} = X > \beta$. Conditioned on that, $\tilde{X}$ has distribution as

follows:

$$f_{\tilde{X}}(x) = \begin{cases} \lambda e^{-\lambda(x-\beta)}, & \text{if } x \geq \beta \\ 0, & \text{if } x < \beta \end{cases}$$

Hence $\tilde{X}$ can be written as

$$\tilde{X} = X_e + \beta$$

where X_e has exponential distribution with parameter λ. Since X and V are independent, $X > \beta$ does not affect the distribution of V. Therefore V has gamma distribution with parameters (k, θ). Moreover, the power constraint P_T satisfies $\theta = \sqrt{P_T}$. Hence by applying Lemmas 2 and 3, the optimal encoding/decoding policies in Case II should be

$$\begin{aligned} Y &= g(\tilde{X}) \\ &= \alpha(\tilde{X} - \beta) - \alpha\lambda^{-1} \\ &= \alpha\tilde{X} - \alpha\beta - \alpha\lambda^{-1} \\ &= \alpha|\tilde{X}| - \alpha\beta - \alpha\lambda^{-1} \\ \hat{X} &= h(\tilde{Y}, S) \\ &= \frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y} + \frac{\gamma}{\gamma+1}\lambda^{-1} + \beta \\ &= S \cdot \left(\frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y} + \frac{\gamma}{\gamma+1}\lambda^{-1} + \beta\right) \end{aligned}$$

Furthermore, the optimal cost is

$$\mathbb{E}[(X - \hat{X})^2 \mid X > \beta] = \frac{1}{\gamma+1}\frac{1}{\lambda^2} = m \tag{4.1}$$

Case III. $U = 1$, $S = -$. Similar to Case II, the problem collapses to the communication problem discussed in chapter 3 with the input signal $\tilde{X}$ and the communication channel noise V. $U = 1$, $S = -$ is equivalent to $\tilde{X} = X < -\beta$. Conditioned on that, $\tilde{X}$ can be written as

$$\tilde{X} = -X_e - \beta$$

where X_e has exponential distribution with parameter λ. V still has gamma distribution with parameters (k, θ). Applying Lemmas 2 and 3 again, the

optimal encoding/decoding policies in Case III should be

$$\begin{aligned} Y &= g(\tilde{X}) \\ &= -\alpha\tilde{X} - \alpha\beta - \alpha\lambda^{-1} \\ &= \alpha|\tilde{X}| - \alpha\beta - \alpha\lambda^{-1} \\ \hat{X} &= h(\tilde{Y}) \\ &= -\frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y} - \frac{\gamma}{\gamma+1}\lambda^{-1} - \beta \\ &= S\cdot\left(\frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y} + \frac{\gamma}{\gamma+1}\lambda^{-1} + \beta\right) \end{aligned}$$

Moreover, the optimal cost is

$$\mathbb{E}[(X-\hat{X})^2 \,\big|\, X < -\beta] = \frac{1}{\gamma+1}\frac{1}{\lambda^2} = m \tag{4.2}$$

□

Theorem 4. *Consider the sensor scheduling and remote estimation problem formulated in chapter 2. Suppose that the sensor is restricted to apply the symmetric threshold-based scheduling policy f with threshold β, $\beta \in (0,\infty)$, and the encoder/decoder apply the corresponding optimal encoding/decoding policies (g,h) described in Theorem 3. Then there exists a unique threshold β^* minimizing the cost function among all the thresholds. Furthermore, $\beta^* = \sqrt{c+m}$, $m = \frac{1}{\gamma+1}\frac{1}{\lambda^2}$.*

Proof. The cost function subject to f with threshold β, g, and h can be written as

$$\begin{aligned} J(f,g,h) &= \mathbb{E}\left[cU + (X-\hat{X})^2\right] \\ &= \mathbb{E}\left[cU + (X-\hat{X})^2 \,\big|\, |X| \le \beta\right]\cdot\mathbb{P}(|X|\le\beta) \\ &+ \mathbb{E}\left[cU + (X-\hat{X})^2 \,\big|\, X > \beta\right]\cdot\mathbb{P}(X>\beta) \\ &+ \mathbb{E}\left[cU + (X-\hat{X})^2 \,\big|\, X < -\beta\right]\cdot\mathbb{P}(X<-\beta) \end{aligned}$$

Consider the expectation in the first term:

$$\begin{aligned} & \mathbb{E}\left[cU + (X - \hat{X})^2 \mid |X| \leq \beta\right] \\ = \; & \mathbb{E}\left[0 + (X - 0)^2 \mid |X| \leq \beta\right] \\ = \; & \mathbb{E}\left[X^2 \mid |X| \leq \beta\right] \\ = \; & \int_{-\beta}^{\beta} x^2 f_X(x) \frac{1}{\mathbb{P}(|X_t| < \beta)} dx \end{aligned}$$

Now consider the expectation in the second term:

$$\begin{aligned} & \mathbb{E}\left[cU + (X - \hat{X})^2 \mid X > \beta\right] \\ = \; & \mathbb{E}\left[c + (X - \hat{X})^2 \mid X > \beta\right] \\ = \; & c + \mathbb{E}\left[(X - \hat{X})^2 \mid X > \beta\right] \\ = \; & c + m \end{aligned}$$

where the last equality is due to (4.1). Similarly, by equation (4.2), the expectation in the third term is

$$\mathbb{E}\left[cU + (X - \hat{X})^2 \mid X < -\beta\right] = c + m$$

Hence,

$$\begin{aligned} J(f, g, h) &= \int_{-\beta}^{\beta} x^2 f_X(x) dx + (c + m) \int_{\beta}^{\infty} f_X(x) dx + (c + m) \int_{-\infty}^{-\beta} f_X(x) dx \\ &= 2 \int_{0}^{\beta} x^2 f_X(x) dx + 2(c + m) \int_{\beta}^{\infty} f_X(x) dx \end{aligned}$$

Taking derivative of $J(f, g, h)$ with respect to β,

$$\frac{d}{d\beta} J(f, g, h) = 2\beta^2 f_X(\beta) - 2(c + m) f_X(\beta)$$

Since $f_X(\beta) > 0$, we have

$$\frac{d}{d\beta}J(f,g,h) = \begin{cases} 0, & \text{if } \beta = \sqrt{c+m} := \beta^* \\ < 0, & \text{if } \beta \in (0, \beta^*) \\ > 0, & \text{if } \beta \in (\beta^*, \infty) \end{cases}$$

Hence β^* is the unique global minimizer.

□

Remark 2. *Theorem 4 implies that by restricting the sensor scheduling policies to the symmetric threshold-based class, the threshold-based scheduling policy with threshold β^* and the corresponding optimal encoding/decoding policies described in Theorem 3 are globally optimal. Hence, (f, g, h) described in Theorem 3 with threshold β^* is a sub-optimal solution to the problem formulated in chapter 2.*

Theorem 5. *Consider the sensor scheduling and remote estimation problem formulated in chapter 2. The sensor scheduling policy and the encoding/decoding policies (f^*, g^*, h^*) described in Theorem 3 with threshold $\beta^* = \sqrt{c+m}$ are person-by-person optimal.*

Proof. The fact that (g^*, h^*) are jointly optimal encoding/decoding policies corresponding to f^* has been proved in Theorem 3. So we only need to prove the other direction.

When $X = x \geq 0$, the cost of no communication is

$$J_{nc}(x) = x^2$$

while the cost of communication is

$$
\begin{aligned}
& J_c(x) \\
= \;& c + \mathbb{E}\left[\left(x - \Big(\frac{1}{\alpha}\frac{\gamma}{\gamma+1}\tilde{Y} + \frac{\gamma}{\gamma+1}\lambda^{-1} + \beta^*\Big)\right)^2\right] \\
= \;& c + \mathbb{E}\left[\left(x - \Big(\frac{1}{\alpha}\frac{\gamma}{\gamma+1}(Y+V) + \frac{\gamma}{\gamma+1}\lambda^{-1} + \beta^*\Big)\right)^2\right] \\
= \;& c + \mathbb{E}\left[\left(x - \Big(\frac{1}{\alpha}\frac{\gamma}{\gamma+1}(\alpha x - \alpha\beta^* - \alpha\lambda^{-1} + V) + \frac{\gamma}{\gamma+1}\lambda^{-1} + \beta^*\Big)\right)^2\right] \\
= \;& c + \mathbb{E}\left[\left(x - \Big(\frac{\gamma}{\gamma+1}x + \frac{1}{\gamma+1}\beta^* + \frac{1}{\alpha}\frac{\gamma}{\gamma+1}V\Big)\right)^2\right] \\
= \;& c + \mathbb{E}\left[\left(\frac{1}{\gamma+1}\big(x - \beta^*\big) - \frac{1}{\alpha}\frac{\gamma}{\gamma+1}V\right)^2\right] \\
= \;& c + \frac{1}{(\gamma+1)^2}\big(x-\beta^*\big)^2 - \frac{2\gamma}{\alpha(\gamma+1)^2}\big(x-\beta^*\big)\mathbb{E}[V] + \frac{1}{\alpha^2}\frac{\gamma^2}{(\gamma+1)^2}\mathbb{E}[V^2] \\
= \;& c + \frac{1}{(\gamma+1)^2}\big(x-\beta^*\big)^2 - \frac{2\gamma}{\alpha(\gamma+1)^2}\big(x-\beta^*\big)\mathbb{E}[V] + \frac{1}{\alpha^2}\frac{\gamma^2}{(\gamma+1)^2}\mathbb{E}[V]^2 \\
& + \frac{1}{\alpha^2}\frac{\gamma^2}{(\gamma+1)^2}\mathrm{Var}(V) \\
= \;& c + \left(\frac{1}{\gamma+1}\big(x-\beta^*\big) - \frac{1}{\alpha}\frac{\gamma}{\gamma+1}\mathbb{E}[V]\right)^2 + \frac{1}{\alpha^2}\frac{\gamma^2}{(\gamma+1)^2}\mathrm{Var}(V) \\
= \;& c + \left(\frac{1}{\gamma+1}\big(x-\beta^*\big) - \frac{1}{\alpha}\frac{\gamma}{\gamma+1}k\theta\right)^2 + \frac{1}{\alpha^2}\frac{\gamma^2}{(\gamma+1)^2}k\theta^2 \\
= \;& c + \frac{1}{(\gamma+1)^2}\big(x-\beta^*-\alpha^{-1}\gamma k\theta\big)^2 + \frac{1}{\alpha^2}\frac{\gamma^2}{(\gamma+1)^2}k\theta^2
\end{aligned}
$$

The expression above can be simplified by applying the assumptions on parameters (2.1). Hence,

$$J_c(x) = c + \frac{1}{(\gamma+1)^2}\big(x-\beta^*-\lambda^{-1}\big)^2 + \frac{\gamma}{(\gamma+1)^2}\lambda^{-2}$$

First one should see that

$$\begin{aligned} J_c(\beta^*) &= c + \frac{1}{(\gamma+1)^2}\,\lambda^{-2} + \frac{\gamma}{(\gamma+1)^2}\,\lambda^{-2} \\ &= c + \frac{1}{\gamma+1}\frac{1}{\lambda^2} \\ &= c + m \\ &= (\beta^*)^2 \\ &= J_{nc}(\beta^*) \end{aligned}$$

When $x \in [0, \beta^*)$,

$$\begin{aligned} \frac{d}{dx}J_{nc}(x) &= 2x > 0 \\ \frac{d}{dx}J_c(x) &= \frac{2}{(\gamma+1)^2}(x - \beta^* - \lambda^{-1}) < 0 \end{aligned}$$

which implies

$$J_{nc}(x) < J_{nc}(\beta^*) = J_c(\beta^*) < J_c(x) \;\forall\; x \in [0, \beta^*)$$

When $x \in (\beta^*, \infty)$,

$$\begin{aligned} \frac{d}{dx}\Big(J_{nc}(x) - J_c(x)\Big) &= \frac{d}{dx}J_{nc}(x) - \frac{d}{dx}J_c(x) \\ &= 2x - \frac{2}{(\gamma+1)^2}(x - \beta^* - \lambda^{-1}) \\ &= \frac{2(\gamma^2+2\gamma)}{(\gamma+1)^2}\,x + \frac{2}{(\gamma+1)^2}(\beta^* + \lambda^{-1}) \\ &> 0 \end{aligned}$$

which implies

$$J_{nc}(x) - J_c(x) > 0, \quad \forall\; x \in (\beta^*, \infty)$$

Therefore

$$\begin{aligned} U = f^*(X) &= 0, \qquad \text{if } X \in [0, \beta^*] \\ U = f^*(X) &= 1, \qquad \text{if } X \in (\beta^*, \infty) \end{aligned}$$

By symmetry one can also obtain that

$$U = f^*(X) = 0, \qquad \text{if } X \in [-\beta^*, 0]$$
$$U = f^*(X) = 1, \qquad \text{if } X \in (-\infty, -\beta^*)$$

□

Remark 3. *Consider the asymptotic case where $P_T \to \infty$ and $k\theta, k\theta^2 \to 0$ (which means $\mathbb{E}[V], \mathbb{E}[V^2] \to 0$). Although $\theta = \sqrt{P_T} \to \infty$, k can be P_T^{-2} such that $k\theta, k\theta^2 \to 0$. Then $\gamma = k^{-1} \to \infty$, which implies that $m = \frac{1}{\gamma+1}\frac{1}{\lambda^2} \to 0$. Hence the optimal sensor scheduling policy becomes*

$$U = \begin{cases} 1, & \text{if } |X| > \beta^* \\ 0, & \text{if } |X| \leq \beta^* \end{cases}$$

where $\beta^* = \sqrt{c}$. Furthermore, since $\alpha = \lambda\sqrt{P_T} \to \infty$ and $\mathbb{E}[V], \mathbb{E}[V^2] \to 0$, we have $\frac{1}{\alpha}\tilde{Y} = \frac{1}{\alpha}(Y + V) \to \frac{1}{\alpha}Y$, and the output of the decoder becomes

$$\hat{X} = \begin{cases} X, & \text{if } U = 1 \\ 0, & \text{if } U = 0 \end{cases}$$

Note that this result matches the results by Nayyar et al [29], which have been shown to be globally optimal. In that paper the source was assumed to have a Gaussian distribution, while the results and the proof also work for a source with Laplace distribution, which is symmetric and unimodal.

4.3 Results on a Modified Problem

4.3.1 Reformulation of the Problem

We first allow the sensor to directly map X to $\tilde{X}$, namely

$$\tilde{X} = f(X)$$

where f can be an arbitrary function mapping $\mathbb{R}$ to $\mathbb{R}$. $\tilde{X} = 0$ means the sensor decides not to transmit. Define the quantizing function $Q(\tilde{X})$ as

follows:

$$Q(\tilde{X}) = \begin{cases} +1 & , \text{ if } \tilde{X} > 0 \\ -1 & , \text{ if } \tilde{X} < 0 \\ 0 & , \text{ if } \tilde{X} = 0 \end{cases}$$

Then the communication cost is $c \cdot (Q(\tilde{X}))^2$. Once the encoder receives the message $\tilde{X}$ from the sensor, it sends the encoded message Y to the communication channel. Hence,

$$Y = g(\tilde{X})$$

In particular, the encoder sends $Y = 0$ if the received message $\tilde{X} = 0$. Meanwhile, the encoder sends $Q(\tilde{X})$ to the decoder via the side channel. The encoder has power constraint, namely

$$\mathbb{E}\left[Y^2\right] \leq P_T'$$

The decoder generates an estimate of X, denoted by $\hat{X}$, based on the messages $\tilde{Y}$ and $Q(\tilde{X})$, namely

$$\hat{X} = h(\tilde{Y}, Q(\tilde{X}))$$

where $\tilde{Y} = Y + V$. The estimation cost is the squared error $(X - \hat{X})^2$.

We still assume that X has exponential distribution with parameter λ, and V has gamma distribution with parameters (k, θ). Furthermore, assume that

$$\begin{aligned} \theta &= \sqrt{P_T} \\ P_T' &= e^{-\lambda\beta^*} P_T \end{aligned}$$

where $\beta^* = \sqrt{c+m}, m = \frac{1}{\gamma+1}\frac{1}{\lambda^2}$. Then

$$\begin{aligned} \alpha &= \lambda\sqrt{P_T} = \lambda\theta \\ \gamma &= \frac{P_T}{k\theta^2} = \frac{1}{k} \end{aligned}$$

The objective is to design (f, g, h) to minimize the cost function

$$J(f, g, h) = \mathbb{E}\left[c \cdot (Q(\tilde{X}))^2 + (X - \hat{X})^2\right]$$

subject to the power constraint of the encoder.

4.3.2 Local Optimality of the Proposed Solution

Lemma 4. *Consider the reformulated problem described in section 4.3.1. If the sensor applies scheduling policy f^* as follows:*

$$\tilde{X} = f^*(X) = \begin{cases} X, & \text{if } |X| > \beta^* \\ 0, & \text{if } |X| \leq \beta^* \end{cases}$$

where $\beta^* = \sqrt{c+m}$,$m = \frac{1}{\gamma+1}\frac{1}{\lambda^2}$, then the encoding/decoding policies (g^*, h^*) described below are jointly optimal among all the policies satisfying the power constraint:

$$\begin{aligned} g^*(\tilde{X}) &= \begin{cases} \alpha|\tilde{X}| - \alpha\beta^* - \alpha\lambda^{-1}, & \text{if } \tilde{X} \neq 0 \\ 0, & \text{if } \tilde{X} = 0 \end{cases} \\ h^*(\tilde{Y}, Q(\tilde{X})) &= \begin{cases} Q(\tilde{X}) \cdot \left(\frac{1}{\alpha} \frac{\gamma}{\gamma+1} \tilde{Y} + \frac{\gamma}{\gamma+1} \lambda^{-1} + \beta^* \right), & \text{if } Q(\tilde{X}) \neq 0 \\ 0, & \text{if } Q(\tilde{X}) = 0 \end{cases} \end{aligned}$$

where $\alpha = \lambda\sqrt{P_T}$, $\gamma = \frac{P_T}{k\theta^2}$

Proof. First observe that

$$\begin{aligned} \mathbb{E}[Y^2] &= \mathbb{E}\left[Y^2 \mid \tilde{X} \neq 0\right] \cdot \mathbb{P}(\tilde{X} \neq 0) + \mathbb{E}\left[Y^2 \mid \tilde{X} = 0\right] \cdot \mathbb{P}(\tilde{X} = 0) \\ &= \mathbb{E}\left[Y^2 \mid \tilde{X} \neq 0\right] \cdot \mathbb{P}(\tilde{X} \neq 0) + 0 \cdot \mathbb{P}(\tilde{X} = 0) \\ &= \mathbb{E}\left[Y^2 \mid \tilde{X} \neq 0\right] \cdot \mathbb{P}(\tilde{X} \neq 0) \\ &\leq P_T' \\ &= e^{-\lambda\beta^*} P_T \end{aligned}$$

If the sensor applies the scheduling policy f^*, then

$$\mathbb{P}(\tilde{X} \neq 0) = \mathbb{P}(|X| \leq \beta^*) = e^{-\lambda\beta^*}$$

Hence the problem becomes an encoding/decoding problem with power constraint

$$\mathbb{E}\left[Y^2 \mid \tilde{X} \neq 0\right] \leq P_T$$

which has been solved in Theorem 3. Hence the optimal encoding/decoding policies corresponding to f^* are (g^*, h^*) as described above.

□

Before discussing the local optimality of the proposed solution, we need to define the neighborhood of the policies (f^*, g^*, h^*). Define the norm on f as follows:

$$\|f\| = \sup_{x\in\mathbb{R}} |f(x)| + \sup_{x\in\mathbb{R}} |f'(x)| + \sup_{x\in\mathbb{R}} |Q(f(x))|$$

Then the difference of two functions f^* and f is

$$\|f^* - f\| = \sup_{x\in\mathbb{R}} |f^*(x) - f(x)| + \sup_{x\in\mathbb{R}} |f^{*\prime}(x) - f'(x)| + \sup_{x\in\mathbb{R}} |Q(f^*(x)) - Q(f(x))|$$

Hence we define the neighborhood of f^*, denoted by $\mathcal{N}(f^*)$, as

$$\mathcal{N}(f^*) = \left\{f \mid \|f^* - f\| \leq \delta\right\}$$

where $\delta > 0, \delta << 1$. Then the functions in $\mathcal{N}(f^*)$ can also be expressed in the form of $f^* + \epsilon\delta_1$, where $|\epsilon|$ is sufficiently small and the admissible perturbation function δ_1 satisfies

$$\begin{aligned}
&\delta_1(x) \text{ is differentiable almost everywhere} \\
&\sup_{x\in\mathbb{R}} |\delta_1(x)| < \infty, \quad \sup_{x\in\mathbb{R}} |\delta_1'(x)| < \infty \\
&\delta_1(x) = 0 \text{ almost everywhere when} f(x) = 0
\end{aligned}$$

Similarly, define the norm on g and h as

$$\begin{aligned}
\|g\| &= \sup_{x\in\mathbb{R}} |g(x)| + \sup_{x\in\mathbb{R}} |g'(x)| \\
\|h\| &= \max_{q\in\{\pm 1,0\}} \sup_{x\in\mathbb{R}} |h(x,q)| + \max_{q\in\{\pm 1,0\}} \sup_{x\in\mathbb{R}} \left|\frac{\partial h(x,q)}{\partial x}\right|
\end{aligned}$$

Then the neighborhoods of g^* and h^*, denoted by $\mathcal{N}(g^*)$ and $\mathcal{N}(h^*)$, respec-

tively, are defined as

$$\mathcal{N}(g^*) = \big\{ g \;\big|\; \|g^* - g\| \le \delta \big\}$$
$$\mathcal{N}(h^*) = \big\{ h \;\big|\; \|h^* - h\| \le \delta \big\}$$

where $\delta > 0, \delta << 1$. Therefore, the functions in $\mathcal{N}(g^*)$ and $\mathcal{N}(h^*)$ can be expressed in the form of $g^* + \epsilon\delta_2$ and $h^* + \epsilon\delta_3$, respectively, where $|\epsilon|$ is sufficiently small and the admissible perturbation functions δ_2 and δ_3 satisfy

$$\begin{aligned}
&\delta_2(x), \delta_3(x,q) \text{ are differentiable almost everywhere, } \forall\, q \in \{0, \pm 1\} \\
&\sup_{x\in\mathbb{R}} |\delta_2(x)| < \infty, \quad \sup_{x\in\mathbb{R}} |\delta_2'(x)| < \infty \\
&\sup_{x\in\mathbb{R}} |\delta_3(x,q)| < \infty, \quad \sup_{x\in\mathbb{R}} \left| \frac{\partial \delta_3(x,q)}{\partial x} \right| < \infty, \;\; \forall\, q \in \{0, \pm 1\}
\end{aligned}$$

Based on the neighborhoods defined above, we have the following results:

Theorem 6. *Consider the reformulated problem described in section 4.3.1. The sensor scheduling policy, and the encoding/decoding policies (f^*, g^*, h^*) described in Lemma 4 are locally optimal within the neighborhoods $\mathcal{N}(f^*)$, $\mathcal{N}(g^*)$, $\mathcal{N}(h^*)$ described above.*

Proof. Since the power constraint P_T' is fixed, it is equivalent to solve the unconstrained problem with an additional power cost $\lambda_L \cdot \mathbb{E}[Y^2]$ in the cost function with some Lagrange multiplier λ_L. That is, to minimize the following cost function:

$$\begin{aligned}
& J(f,g,h) \\
= \;& \mathbb{E}\left[c \cdot (Q(\tilde{X}))^2 + \lambda_L \cdot Y^2 + (X - \hat{X})^2 \right] \\
= \;& c \cdot \mathbb{E}\left[(Q(\tilde{X}))^2 \right] + \lambda_L \cdot \mathbb{E}\left[Y^2 \right] + \mathbb{E}\left[(X - \hat{X})^2 \right] \\
= \;& c \cdot \int_{-\infty}^{+\infty} Q^2\big(f(x)\big) f_X(x) dx + \lambda_L \cdot \int_{-\infty}^{+\infty} g^2\big(f(x)\big) f_X(x) dx \\
& + \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \Big(h\big(g(f(x)) + v, Q(f(x))\big) - x \Big)^2 f_X(x) f_V(v) dv dx
\end{aligned}$$

By Lemma 4, if f^* is fixed, then g^* and h^* are globally optimal among all the encoding/decoding policies. Then for any admissible perturbation functions δ_2, δ_3, the first and the second order Gateaux differentials of J

at (f^*, g^*, h^*) with increments δ_2, δ_3, denoted by $\delta J(f^*, g^*, h^*; \delta_2, \delta_3)$ and $\delta^2 J(f^*, g^*, h^*; \delta_2, \delta_3)$, respectively, should satisfy

$$\begin{aligned} \delta J(f^*, g^*, h^*; \delta_2, \delta_3) &= 0 \\ \delta^2 J(f^*, g^*, h^*; \delta_2, \delta_3) &\geq 0 \end{aligned}$$

First let us compute $\delta J(f^*, g^*, h^*; \delta_2, \delta_3)$:

$$\begin{aligned} & \delta J(f^*, g^*, h^*; \delta_2, \delta_3) \\ = \; & \frac{d}{d\epsilon} J(f^*, g^* + \epsilon\delta_2, h^* + \epsilon\delta_3)\Big|_{\epsilon=0} \\ = \; & c \cdot \int_{-\infty}^{+\infty} \frac{d}{d\epsilon} Q^2\big(f^*(x)\big)\Big|_{\epsilon=0} f_X(x)dx \\ & + \lambda_L \cdot \int_{-\infty}^{+\infty} \frac{d}{d\epsilon}\Big(g^*\big(f^*(x)\big) + \epsilon\delta_2\big(f^*(x)\big)\Big)^2\Big|_{\epsilon=0} f_X(x)dx \\ & + \int_{-\infty}^{+\infty}\int_{0}^{+\infty} \frac{d}{d\epsilon}\Bigg(h^*\Big(g^*\big(f^*(x)\big) + \epsilon\delta_2\big(f^*(x)\big) + v, Q\big(f^*(x)\big)\Big) \\ & \quad + \epsilon\delta_3\Big(g^*\big(f^*(x)\big) + \epsilon\delta_2\big(f^*(x)\big) + v, Q\big(f^*(x)\big)\Big) - x\Bigg)^2\Bigg|_{\epsilon=0} \\ & \quad f_X(x) f_V(v) dv dx \end{aligned}$$

Consider the first term,

$$\frac{d}{d\epsilon} Q^2\big(f^*(x)\big)\Big|_{\epsilon=0} = 0$$

Consider the second term,

$$\begin{aligned} & \frac{d}{d\epsilon}\Big(g^*\big(f^*(x)\big) + \epsilon\delta_2\big(f^*(x)\big)\Big)^2\Big|_{\epsilon=0} \\ = \; & g^*\big(f^*(x)\big) \cdot \delta_2\big(f^*(x)\big) \\ = \; & \begin{cases} 0, & \text{if } |x| \leq \beta^* \\ g^*(x) \cdot \delta_2\big(x\big), & \text{if } |x| > \beta^* \end{cases} \end{aligned}$$

Consider the third term,

$$
\begin{aligned}
& \frac{d}{d\epsilon}\Big(h^*\Big(g^*\big(f^*(x)\big) + \epsilon\delta_2\big(f^*(x)\big) + v, Q\big(f^*(x)\big)\Big) \\
& +\epsilon\delta_3\Big(g^*\big(f^*(x)\big) + \epsilon\delta_2\big(f^*(x)\big) + v, Q\big(f^*(x)\big)\Big) - x\Big)^2\Big|_{\epsilon=0} \\
= \;& \Big(h^*\Big(g^*\big(f^*(x)\big) + v, Q\big(f^*(x)\big)\Big) - x\Big) \cdot \Big(\frac{\partial h^*}{\partial \tilde{y}} \cdot \delta_2\big(f^*(x)\big) \\
& + \delta_3\Big(g^*\big(f^*(x)\big) + v, Q\big(f^*(x)\big)\Big)\Big) \\
= \;& \begin{cases} -x \cdot \delta_3(v,0), & \text{if } |x| \le \beta^* \\ \Big(h^*\big(g^*(x)+v,1\big) - x\Big) \cdot \Big(\frac{1}{\alpha}\frac{\gamma}{\gamma+1} \cdot \delta_2\big(x\big) & \\ \quad +\delta_3\Big(g^*\big(x\big)+v,1\Big)\Big), & \text{if } x > \beta^* \\ \Big(h^*\big(g^*(x)+v,-1\big) - x\Big) \cdot \Big(-\frac{1}{\alpha}\frac{\gamma}{\gamma+1} \cdot \delta_2\big(x\big) & \\ \quad +\delta_3\Big(g^*\big(x\big)+v,-1\Big)\Big), & \text{if } x < -\beta^* \end{cases}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& \delta J(f^*, g^*, h^*; \delta_2, \delta_3) \\
= \;& \lambda_L \cdot \int_{-\infty}^{-\beta^*} g^*(x) \cdot \delta_2\big(x\big) f_X(x) dx + \lambda_L \cdot \int_{\beta^*}^{\infty} g^*(x) \cdot \delta_2\big(x\big) f_X(x) dx \\
& + \int_0^{\infty}\int_{\beta^*}^{\infty} \Big(h^*\big(g^*(x)+v,1\big) - x\Big) \\
& \quad \cdot \Big(\frac{1}{\alpha}\frac{\gamma}{\gamma+1} \cdot \delta_2\big(x\big) + \delta_3\Big(g^*\big(x\big)+v,1\Big)\Big) f_X(x) f_V(v) dx dv \\
& + \int_0^{\infty}\int_{-\infty}^{-\beta^*} \Big(h^*\big(g^*(x)+v,-1\big) - x\Big) \\
& \quad \cdot \Big(-\frac{1}{\alpha}\frac{\gamma}{\gamma+1} \cdot \delta_2\big(x\big) + \delta_3\Big(g^*\big(x\big)+v,-1\Big)\Big) f_X(x) f_V(v) dx dv \\
= \;& 0 \quad \text{for all admissible } \delta_2,\ \delta_3
\end{aligned}
$$

Similarly, we can calculate $\delta J^2(f^*, g^*, h^*; \delta_2, \delta_3)$ as follows:

$$
\begin{aligned}
& \delta J^2(f^*, g^*, h^*; \delta_2, \delta_3) \\
= \; & \lambda_L \cdot \int_{-\infty}^{-\beta^*} \Big(\delta_2\big(x\big)\Big)^2 f_X(x)dx + \lambda_L \cdot \int_{\beta^*}^{\infty} \Big(\delta_2\big(x\big)\Big)^2 f_X(x)dx \\
& + \int_0^{\infty} \int_{\beta^*}^{\infty} \left(\frac{1}{\alpha}\frac{\gamma}{\gamma+1} \cdot \delta_2\big(x\big) + \delta_3\Big(g^*\big(x\big) + v, 1\Big)\right)^2 f_X(x) f_V(v) dxdv \\
& + \int_0^{\infty} \int_{-\infty}^{-\beta^*} \left(-\frac{1}{\alpha}\frac{\gamma}{\gamma+1} \cdot \delta_2\big(x\big) + \delta_3\Big(g^*\big(x\big) + v, -1\Big)\right)^2 f_X(x) f_V(v) dxdv \\
& + \int_0^{\infty} \int_{\beta^*}^{\infty} \Big(h^*\big(g^*(x) + v, 1\big) - x\Big) \cdot \frac{\partial \delta_3(\tilde{y}, 1)}{\partial \tilde{y}} \cdot \delta_2\big(x\big) \cdot f_X(x) f_V(v) dxdv \\
& + \int_0^{\infty} \int_{-\infty}^{-\beta^*} \Big(h^*\big(g^*(x) + v, -1\big) - x\Big) \cdot \frac{\partial \delta_3(\tilde{y}, -1)}{\partial \tilde{y}} \\
& \quad \cdot \delta_2\big(x\big) \cdot f_X(x) f_V(v) dxdv \\
\geq \; & 0 \text{ for all admissible } \delta_2,\ \delta_3
\end{aligned}
$$

Now we can compute the first and the second order Gateaux differentials of J at (f^*, g^*, h^*) with increments δ_1, δ_2, δ_3, denoted by $\delta J(f^*, g^*, h^*; \delta_1, \delta_2, \delta_3)$ and $\delta^2 J(f^*, g^*, h^*; \delta_1, \delta_2, \delta_3)$, respectively. According to the properties of the admissible perturbation function δ_1, when $|\epsilon|$ is small enough,

$$Q\big(f^*(x) + \epsilon\delta_1(x)\big) = Q\big(f^*(x)\big) \text{ almost everywhere}$$

Following the same process when computing $\delta J(f^*, g^*, h^*; \delta_2, \delta_3)$, one can get

$$
\begin{aligned}
& \delta J(f^*, g^*, h^*; \delta_1, \delta_2, \delta_3) \\
= \; & \frac{d}{d\epsilon} J(f^* + \epsilon\delta_1, g^* + \epsilon\delta_2, h^* + \epsilon\delta_3)\Big|_{\epsilon=0} \\
= \; & \delta J(f^*, g^*, h^*; \tilde{\delta}_2, \delta_3)
\end{aligned}
$$

where

$$
\tilde{\delta}_2(x) = \begin{cases} \delta_2(x), & \text{if } |x| \leq \beta^* \\ \delta_2(x) + \alpha\delta_1(x) & \text{if } x > \beta^* \\ \delta_2(x) - \alpha\delta_1(x) & \text{if } x < -\beta^* \end{cases}
$$

By the properties of $\delta_1(x)$ and $\delta_2(x)$, we have

$$\tilde{\delta}_2(x) \text{ is differentiable almost everywhere}$$
$$\sup_{x\in\mathbb{R}} |\tilde{\delta}_2(x)| < \infty, \quad \sup_{x\in\mathbb{R}} |\tilde{\delta}_2'(x)| < \infty$$

which means $\tilde{\delta}_2(x)$ is an admissible perturbation function for g^*. Hence

$$\delta J(f^*, g^*, h^*; \delta_1, \delta_2, \delta_3) = 0 \quad \text{for all admissible } \delta_1,\ \delta_2,\ \delta_3$$

Similarly,

$$\begin{aligned} & \delta J^2(f^*, g^*, h^*; \delta_1, \delta_2, \delta_3) \\ = \ & \delta J^2(f^*, g^*, h^*; \tilde{\delta}_2, \delta_3) \\ \geq \ & 0 \quad \text{for all admissible } \delta_1,\ \delta_2,\ \delta_3 \end{aligned}$$

Hence (f^*, g^*, h^*) is a local optimal solution.

□

4.3.3 Existence of Globally Optimal Solution

Theorem 7. *The reformulated problem described in section 4.3.1 has a globally optimal solution.*

Proof. We first relax the decision policies of the sensor, the encoder, and the decoder from deterministic functions

$$\begin{aligned} \tilde{X} &= f(X) \\ Y &= g(\tilde{X}) \\ \hat{X} &= h(\tilde{Y}, Q(\tilde{X})) \end{aligned}$$

to probabilistic transformations, expressed as $f_{\tilde{X}|X}, f_{Y|\tilde{X}}$ and $f_{\hat{X}|\tilde{Y},Q(\tilde{X})}$. We define the *generalized* problem as: minimize the cost function over the conditional densities $f_{\tilde{X}|X}, f_{Y|\tilde{X}}$ and $f_{\hat{X}|\tilde{Y},Q(\tilde{X})}$. Since the cost function is continuous on $f_{\tilde{X}|X}, f_{Y|\tilde{X}}$ and $f_{\hat{X}|\tilde{Y},Q(\tilde{X})}$, and the sets of probabilistic transformations are compact, the optimal solution of the *generalized* problem exists.

We now show that the minimum of the cost function in the generalized problem can be achieved by deterministic policies. First, we observe that

optimal $f_{\hat{X}|\tilde{Y},Q(\tilde{X})}$ is deterministic since the deterministic function

$$h(\tilde{Y}, Q(\tilde{X})) = \mathbb{E}\left[X \mid \tilde{Y}, Q(\tilde{X})\right]$$

minimizes the first term $\mathbb{E}\left[(X - \hat{X})^2\right]$ and other terms are not affected by this choice of $h(\cdot, \cdot)$. Next, for any deterministic $h(\cdot, \cdot)$, the cost function can be written as

$$\int f_X(x) \left(\int G_V(x, y, \tilde{x}) f_{Y|\tilde{X}}(y|\tilde{x}) f_{\tilde{X}|X}(\tilde{x}|x) dy d\tilde{x} \right) dx \tag{4.3}$$

where

$$\begin{aligned} & G_V(x, y, \tilde{x}) \\ &= \int \left(\left(x - h(y + v, Q(x))\right)^2 + c \cdot Q^2(f(x)) + \lambda_L \cdot y^2 \right) f_V(v) dv \end{aligned}$$

The minimization in (4.3) can be carried out by choosing the $\tilde{X} = \tilde{x}, Y = y$ pair that minimizes $G_V(x, y, \tilde{x})$ for each $X = x$. Using the optimal $h(\cdot, \cdot)$ as the fixed $h(\cdot, \cdot)$ in (4.3), it follows that the optimal $\tilde{X}$ is a deterministic function of X and optimal Y is a deterministic function of $\tilde{X}$. Therefore, the optimal solutions of the *generalized* problem and the original problem are the same. Then the optimal solution of the original problem also exists.

□

CHAPTER 5

CONCLUSIONS

In this thesis, we considered a sensor scheduling and remote estimation problem with a noisy communication channel between the sensor and the estimator. Our goal was to design the communication strategy and the encoding strategy for the sensor, and the decoding strategy for the estimator, that minimize the expected value of the sum of communication cost and estimation cost over a finite time horizon. By making i.i.d. assumptions on the source and the noise processes, we showed that one should be able to find the optimal solution among stationary strategies by solving the "one-stage" problem. Furthermore, by assuming that the source and the noise have Laplace and gamma distributions, respectively, with specific parameters, we obtained a person-by-person optimal solution, which consists of a symmetric threshold-based sensor scheduling strategy and a pair of piecewise linear encoding/decoding strategies. The person-by-person optimal solution is shown to be globally optimal in the asymptotic case. Furthermore, in a modified problem it is shown that our solution is locally optimal and a globally optimal solution exists.

Future directions for research include:

(1) Extending the results to other types of distributions for the source and the noise. To conduct this extension one may need to assume that the channel noise is generated by an adversary or jammer, who wants to maximize the estimation error. Then the optimization problem becomes a zero-sum game and we are interested in finding the communication scheduling and encoding strategies, the decoding strategy, and the jamming strategy for the sensor, the estimator, and the jammer, respectively, that achieve a saddle-point equilibrium. Some related works can be found in [36, 37, 38].

(2) Extending the results to Markov source. To conduct this extension one may need to assume that there is a noise-free feedback channel from the estimator to the sensor. By the end of each time step, the estimator

transmits its current estimate to the sensor.

(3) Extending the results to multiple communication channels. Each channel is associated with a power constraint and a communication cost. Therefore, when the sensor decides to communicate with the estimator, it also needs to decide which channel to use. Assume that the estimator knows the channel chosen by the sensor. Then the estimator should also apply different decoding strategies for messages transmitted via different channels.

(4) Extending the results to multi-dimensional systems.

REFERENCES

[1] D. Shuman, A. Nayyar, A. Mahajan, Y. Goykhman, K. Li, M. Liu, D. Teneketzis, M. Moghaddam, and D. Entekhabi, "Measurement scheduling for soil moisture sensing: From physical models to optimal control," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1918–1933, 2010.

[2] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, 2007.

[3] "Industrial wireless technology for the 21st century," based on views of the industrial wireless community, US Department of Energy, 2002.

[4] I. Akyildiz, W. Su, S. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.

[5] M. Athans, "On the determination of optimal costly measurement strategies for linear stochastic systems," *Automatica*, vol. 8, pp. 397–412, 1972.

[6] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, pp. 251–260, 2006.

[7] C. Yang and L. Shi, "Deterministic sensor data scheduling under limited communication resource," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 5050–5056, 2011.

[8] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "Stochastic sensor scheduling for energy constrained estimation in multi-hop wireless sensor networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2489–2495, 2011.

[9] L. Shi and H. Zhang, "Scheduling two Gauss-Markov systems: An optimal solution for remote state estimation under bandwidth constraint," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2038–2042, 2012.

[10] Z. Ren, P. Cheng, J. Chen, L. Shi, and Y. Sun, "Optimal periodic sensor schedule for steady-state estimation under average transmission energy constraint," *IEEE Transactions on Automatic Control*, vol. 58, no. 12, pp. 3265–3271, 2013.

[11] K. J. Astrom and B. M. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," in *Proceedings of the 41st IEEE Conference on Decision and Control*, 2002, pp. 2011–2016.

[12] K. Nar and T. Başar, "Sampling multidimensional Wiener processes," in *Proceedings of the 53rd IEEE Conference on Decision and Control*, 2013, to appear.

[13] O. C. Imer and T. Başar, "Optimal estimation with limited measurements," *International Journal of Systems Control and Communications*, vol. 2, no. 1-3, pp. 5–29, 2010.

[14] O. C. Imer, "Optimal estimation and control under communication network constraints," Ph.D. dissertation, University of Illinois, Urbana-Champaign, 2005.

[15] P. R. Kumar and P. Varaiya, *Stochastic systems: estimation, identification and adaptive control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[16] P. Bommannavar and T. Başar, "Optimal estimation over channels with limits on usage," in *17th IFAC World Congress*, 2008, pp. 6632–6637.

[17] M. Rabi, G. V. Moustakides, and J. S. Baras, "Multiple sampling for real-time estimation on a finite horizon," in *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006, pp. 1351–1357.

[18] M. Rabi, "Packet based inference and control," Ph.D. dissertation, University of Maryland, College Park, 2006.

[19] Y. Xu and J. Hespanha, "Optimal communication logics for networked control systems," in *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2004, pp. 3527–3532.

[20] J. Wu, Q. Jia, K. H. Johansson, and L. Shi, "Event-based sensor data scheduling: Trade-off between communication rate and estimation quality," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 1041–1046, 2013.

[21] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4782–4795, 2006.

[22] E. J. Msechu, S. I. Roumeliotis, A. Ribeiro, and G. B. Giannakis, "Decentralized quantized Kalman filtering with scalable communication cost," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3727–3741, 2008.

[23] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592–2601, 2003.

[24] K. You and L. Xie, "Kalman filtering with scheduled measurements," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1520–1530, 2013.

[25] D. Han, Y. Mo, J. Wu, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor scheduling for remote state estimation," in *Proceedings of the 52nd IEEE Conference on Decision and Control*, 2013, pp. 6079–6084.

[26] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, 2004.

[27] K. You, M. Fu, and L. Xie, "Mean square stability for Kalman filtering with Markovian packet losses," *Automatica*, vol. 47, no. 12, pp. 2647–2657, 2011.

[28] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order LTI systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, 2011.

[29] A. Nayyar, T. Başar, D. Teneketzis, and V. V. Veeravalli, "Optimal strategies for communication and remote estimation with an energy harvesting sensor," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2246–2260, 2013.

[30] B. Hajek, K. Mitzel, and S. Yang, "Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm," *IEEE Transactions on Information Theory*, vol. 54, no. 2, pp. 608–622, 2008.

[31] A. Gupta, C. Langbort, and T. Başar, "Optimal control in the presence of an intelligent jammer with limited actions," in *Proceedings of the 49th IEEE Conference on Decision and Control*, 2010, pp. 1096–1101.

[32] A. Gupta, A. Nayyar, C. Langbort, and T. Başar, "A dynamic transmitter-jammer game with asymmetric information," in *Proceedings of the 51st IEEE Conference on Decision and Control*, 2012, pp. 6477 – 6482.

[33] A. Mahajan and D. Teneketzis, "Optimal design of sequential real-time communication systems," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5317–5338, 2009.

[34] A. Mahajan, "Sequential decomposition of sequential teams: applications to real-time communication and networked control systems," Ph.D. dissertation, University of Michigan, Ann Arbor, 2008.

[35] E. Akyol, K. Viswanatha, K. Rose, and T. Ramstad, "On zero delay source-channel coding," *IEEE Transactions on Information Theory*, accepted, available online at: http://arxiv.org/abs/1302.3660.

[36] T. Başar, "The Gaussian test channel with an intelligent jammer," *IEEE Transactions on Information Theory*, vol. 29, no. 1, pp. 152–157, 1983.

[37] T. Başar and Y. W. Wu, "A complete characterization of minimax and maximin encoder-decoder policies for communication channels with incomplete statistical description," *IEEE Transactions on Information Theory*, vol. 31, no. 4, pp. 482–489, 1985.

[38] E. Akyol, K. Rose, and T. Başar, "On optimal jamming over an additive noise channel," in *Proceedings of the 52nd IEEE Conference on Decision and Control*, 2013, pp. 3079–3084.