Stochastic Packet Scheduling for Optimal Parameter Estimation

Duo Han*, Keyou You†, Lihua Xie*, Junfeng Wu‡, Ling Shi‡

Abstract—In this paper we consider optimal parameter estimation with a constrained packet transmission rate. Due to the limited battery power and the traffic congestion over a large sensor network, each sensor is required to discard some packets and save transmission times. We propose a packet-driven sensor scheduling policy such that the sensor transmits only the important measurements to the estimator. Unlike the existing deterministic scheduler in [1], our stochastic packet scheduling is novelly designed to maintain the computational simplicity of the resulting maximum-likelihood estimator (MLE). This results in a nice feature that the MLE is still able to be recursively computed in a closed form, and the Cramér-Rao lower bound (CRLB) can be explicitly evaluated. Moreover, an optimization problem is formulated and solved to obtain the optimal parameters of the scheduling policy under which the estimation performance is comparable to the standard MLE (with full measurements) even with a moderate transmission rate. Numerical simulations are included to show the effectiveness.

I. INTRODUCTION

With the wireless communication technology penetrating into all aspects of human life, estimation theory faced new challenges in the last few decades [2]–[5]. One of the most encountered problems is how to optimally manage the limited sensor resources such as battery power and network bandwidth [6]–[8]. In this paper, we investigate a class of problems called controlled communication for estimation [9]. The problem reflects a conflict between the sensors and the estimator. The estimator wants as much information as it can obtain to perform estimation while the sensor hopes to work longer since the energy budget is limited and the battery is often irreplaceable for a sparsely distributed sensor network. It is known that the communication unit consumes more power than any other functional block of a sensor [10], [11]. Moreover, a typical network data packet usually consists of many bits of header and thus the cost of transmitting one packet is relatively high no matter how much the payload is.

*: School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Email: dhanaa@ntu.edu.sg, elhxie@nju.edu.sg. The work by L. Xie was partially supported by National Research Foundation of Singapore under NRF2011NRF-CRP001-090 and NRF2013EWT-EIRP004-012, and Natural Science Foundation of China under NSFC 6120106011.
†: Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: eesling@ust.hk. The work by L. Shi is supported by an RGC General Research Fund 16209114.
‡: Department of Automation, Tsinghua University, Beijing, 100084, China. Email: youky@tsinghua.edu.cn. The work by K. You is supported by National High-tech R&D Program of China (863 Program) under grant no. 2013AA040700.
‡: Department of Automatic Control, The Royal Institute of Technology (KTH), Stockholm, Sweden. Email: junfengw@kth.se. The work by J. Wu is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

Therefore, the burden of limited energy can be effectively lightened by controlling the rate of packet transmission. In the sequel we refer to the packet transmission rate as the transmission rate. This motivates us to focus on the parameter estimation under the constrained transmission rate. Specifically, the sensors at each time intentionally discard some unimportant data packets to save their communication resources. A closely related work [1] studied the asymptotically optimal parameter estimation problem with scheduled measurements. They derived the MLE with a subset of measurements and analyzed the asymptotic properties of the estimator. However, their MLE lacks a simple form and is computationally demanding.

In our problem setting (see Fig.1), the sensor observes the vector parameter through an observation matrix with additive Gaussian noise. Then the sensor decides whether to send the measurement $x_k$ to the estimator. The state-dependent or innovation-dependent event-triggered transmission is a prevailing approach to obtain a good tradeoff between the estimation quality and the transmission rate [1], [12]–[15]. In the spirit of [14] which dealt with the MMSE state estimation of a dynamical system, we propose a stochastic scheduling solution to solve the transmission-rate-constrained parameter estimation problem. Unlike [14] we study the asymptotic properties of the proposed estimator under scheduled measurements when the number of sensor measurements tends to infinity. The solution is proved to result in a closed-formed MLE and an adaptive recursive estimator which possesses a main advantage over the deterministic threshold-type mechanism in [1]. Furthermore, we formulate an optimization problem to find the optimal parameters in the scheduling policy. The main contributions of this work are twofold:

1) We design a stochastic scheduling policy to reduce the transmission rate while preserving good accuracy of the parameter estimation. We derive an implementable general form of MLE and a simple recursive estimator in which we introduce two extra degrees of freedom. Moreover, we derived an analytical expression of the Cramér-Rao lower bound (CRLB).

2) We formulate an optimization problem to search the optimal parameters such that the asymptotic error covariance is minimized while the transmission rate is
constrained below an upper bound.

In the next section we formally introduce our mathematical formulation.

II. PROBLEM STATEMENT

Consider a parameter estimation problem of estimating \( \theta \in \mathbb{R}^n \) described as
\[
x_k = H_k^\top \theta + \nu_k, \quad 1 \leq k \leq N
\]
where \( k \) is the time index, \( x_k \in \mathbb{R}^m \) is the vector measurement and \( \nu_k \in \mathbb{R}^m \) is an i.i.d. zero-mean additive Gaussian noise random vector with covariance matrix \( \Sigma \in \mathbb{R}^{m \times m} \). The observation matrix \( H_k \in \mathbb{R}^{n \times m} \) satisfies \( \| H_k \|_\infty < \infty \). The parameter \( \theta \) is unknown and is to be identified.

It is well known that maximum-likelihood (ML) estimation is one of the most widely used statistical method to estimate an unknown parameter. Traditionally, the sensor measures the noisy measurements at each time and all the measurements are perfectly received by the estimator. The maximum-likelihood estimator (MLE) can thus perform estimation based on the information set and give the ML estimate. In order to prolong the lifespan of a sensor due to its limited battery energy, we adopt a packet scheduling approach, i.e., forcing the sensor to transmit only a subset of all the packets. In this way the sensor remains idle for time to time to save energy. In [1], only the measurements which deviate from a predefined threshold too much will be transmitted.

We consider a stochastic packet scheduling policy such that those more important measurements are more likely to be received by the estimator. Note that this policy expands the sensor lifetime in an average sense. Let \( \gamma_k \) be the indicator whether the sensor is authorized to transmit the packet at time \( k \), i.e., permitted when \( \gamma_k = 1 \) and denied when \( \gamma_k = 0 \). Before showing the definition of importance, we introduce some notations. The i.i.d. random variables \( \xi_k \) are uniformly distributed over \([0, 1]\). One sequence of vectors \( \{ \tau_k \in \mathbb{R}^m | 1 \leq k \leq N \} \) and one sequence of positive definite matrices \( \{ \Delta_k \in \mathbb{R}^{m \times m} | 1 \leq k \leq N \} \) are the predefined parameters in the policy. Furthermore, define a shorthand for the following quadratic function of \( x \in \mathbb{R}^m \),
\[
Q_x(y, Z) := -\frac{1}{2}(x - y)\top Z^{-1}(x - y), \quad y \in \mathbb{R}^m, \quad Z \in \mathbb{R}^{m \times m}.
\]
Now we are ready to present the stochastic sensor scheduling policy, given as
\[
\gamma_k = \begin{cases} 
1, & \xi_k \in [\exp( Q_{x_k}(\tau_k, \Delta_k) ), 1], \\
0, & \text{otherwise}.
\end{cases}
\]
(2)

Note that the importance is stochastically assigned, unlike it is deterministically defined in [1]. This special design will facilitate the computation which we will see later. The operation principle of the sensor is described as follows. The sensor collects the observation \( x_k \) and computes the real-valued function \( a = \exp( Q_{x_k}(\tau_k, \Delta_k) ) \) where the design of \( \tau_k \) and \( \Delta_k \) will be discussed later. The sensor obtains a realization \( b \) of the random variable \( \xi_k \) and compares \( a \) and \( b \) to make a transmission decision based on (2). If \( \Delta_i < \Delta_j \) and \( \tau_i = \tau_j \), the sensor is more likely to transmit the \( i \)th observation than the \( j \)th observation.

Remark 2.1: If \( \tau_k = H_k^\top \theta \), the transmission probability will be smaller when \( x_k \) is closer to \( H_k^\top \theta \) which means the noise is small. The idea is that the loss of discarding a measurement far from the true value is larger than that of discarding a measurement close to the true value, because the sensor tends to make a guess of a missing measurement around \( \tau_k \). Since we do not know where \( \theta \) is, we may replace \( \theta \) with the latest estimate of \( \theta \) which we shall discuss later.

We use a continuous random variable \( z_k \) to denote the message received by the estimator at time \( k \). Since the probability measure of \( x_k = 0 \) when \( \gamma_k = 1 \) is 0, the fact that \( z_k = 0 \) represents the event of \( \gamma_k = 0 \) will not affect the analytical results in the sequel. The information set at the estimator side is thus \( \mathcal{Z}_N := \{ z_1, \ldots, z_k, \ldots, z_N \} \), where
\[
z_k := \begin{cases} 
x_k, & \gamma_k = 1, \\
0, & \gamma_k = 0.
\end{cases}
\]

The immediate question is what is the MLE with the information set \( \mathcal{Z}_N \). Is it analytically tractable or implementable in a recursive form? In the next section we will discuss the MLE with the information set \( \mathcal{Z}_N \).

III. MAXIMUM LIKELIHOOD ESTIMATION WITH SCHEDULED MEASUREMENTS

The first step to derive the MLE is to find the joint pdf of \( \mathcal{Z}_N \). To this end, we need some preliminary results. We know that \( x_k \) is independently Gaussian distributed, i.e., \( f_{x_k}(x) := \mathcal{N}(H_k^\top \theta, \Sigma) \). Let us define an auxiliary function for the pdf \( f_{x_k}(x) \) given a pair of \( (\tau_k, \Delta_k) \), \( A_{x_k}(\tau_k, \Delta_k) := \frac{1}{\tau_k} \exp( Q_{x_k}(\tau_k, \Delta_k) ) \) where
\[
\alpha_k := \det( I_n + \Sigma \Delta_k^{-1} )^{-\frac{1}{2}} \exp( Q_{x_k}(\tau_k, \Delta_k + \Sigma) ).
\]
We need the following intermediate result to facilitate the derivation of the MLE. The result can be proved by completion of squares.

Lemma 3.1: For each pair of \( (\tau_k, \Delta_k) \), the product of the auxiliary function and the density function \( f_{x_k}(x) \) is still a Gaussian density function of a transformed random variable \( \tilde{x}_k \), i.e., \( A_{x_k}(\tau_k, \Delta_k) \cdot f_{x_k}(x) = f_{\tilde{x}_k}(x) \), where \( f_{\tilde{x}_k}(x) := \mathcal{N}(\omega_k, \Omega_k) \) and
\[
\omega_k := H_k^\top \theta + \Sigma(\Sigma + \Delta_k)^{-1}(\tau_k - H_k^\top \theta),
\]
\[
\Omega_k = \Sigma - \Sigma(\Sigma + \Delta_k)^{-1}\Sigma.
\]
It is shown that the product of two Gaussian-like function is still a Gaussian-like function. The constant \( \alpha_k \) is a normalized constant to ensure \( f_{\tilde{x}_k}(x) \) to be a Gaussian pdf. Now we have an immediate result on the probability of a measurement being selected to be transmitted.

Lemma 3.2: Given the stochastic sensor scheduling rule in (2), the probability of transmission permission for the sensor at time \( k \) is given by
\[
\Pr\{ \gamma_k = 1 \} = 1 - \alpha_k.
\]
Now the result is obvious as the integral of \( f_\xi(x) \) facilitates the MLE derivation. We are now ready to present \( \nabla \hat{\theta} \). The MLE completes the proof.

Taking the Hessian matrix of \( Z \) we have

\[
\mathbf{R} = \mathbf{R}_\theta = \frac{\partial^2 \log f_\theta(x)}{\partial \theta \partial \theta^\top}.
\]

Then the log-likelihood function is written as

\[
\mathcal{L}(\mathbf{Z}; \theta) = \log f_\mathbf{Z}(\mathbf{z}_1, \ldots, \mathbf{z}_n; \theta) = \sum_{k=1}^{n} f_\mathbf{z}_k(x_\gamma) \Pr\{\gamma_k = 0\}^{1 - x_\gamma}.
\]

For the ease of derivation, we define the following functions of \( \theta \):

\[
g_k(\theta) := \gamma_k H_k \Sigma^{-1} x_k - H_k^\top \theta
\]

\[
+ (1 - \gamma_k) H_k (\Sigma + \Delta_k)^{-1} (\tau_k - H_k^\top \theta),
\]

\[
h_k(\theta) := \gamma_k H_k \Sigma^{-1} H_k^\top + (1 - \gamma_k) H_k (\Sigma + \Delta_k)^{-1} H_k^\top.
\]

We shall see that the resemblance between the threshold in the selection policy (2) and the Gaussian distribution facilitates the MLE derivation. We are now ready to present the MLE under the information set \( \mathcal{Z} \).

\textbf{Theorem 3.3:} Consider the estimation problem in (1) and the scheduling policy in (2). The MLE \( \hat{\theta}_N \) based on the scheduled observations is given as

\[
\hat{\theta}_N = \arg \max_{\theta} l_N(\mathbf{Z}; \theta) = \arg \max_{\theta} \mathcal{L}(\mathbf{Z}; \theta) = \arg \max_{\theta} \sum_{k=1}^{n} g_k(\theta).
\]

\textbf{Proof:} The random variable \( \gamma_k \) is dependent on \( x_k \) and \( \zeta_k \). It is easily shown that

\[
\Pr\{\gamma_k = 1\} = \int_{-\infty}^{+\infty} \int_{\{\gamma_k = 1\}} f_\xi(x) dx d\xi = \int_{-\infty}^{+\infty} (1 - \exp(g_\xi(\tau_k, \Delta_k))) f_\xi(x) dx = 1 - \alpha_k \int_{\mathbb{R}^n} f_\xi(x) dx.
\]

Remark 3.4: The main advantage of the proposed policy is to ensure a simple form of the MLE compared with the existing work [1]. In [1] the MLE with the deterministic scheduler cannot be explicitly written and some algorithms are proposed to numerically find the MLE. The high computational complexity indeed restricts the practical use, especially when the multivariate integration is involved for vector measurements. On the contrary, our scheduling policy and the corresponding MLE can be easily implemented and analyzed.

By invoking the techniques of least squares estimator (LSE) [16], we can write the MLE in a recursive way. Denote \( \hat{\theta}_k := \mathbb{E}[\theta|z_1, \ldots, z_k] \) as the estimate of \( \theta \) after \( k \) measurements and \( P_k := \mathbb{E}[\theta - \hat{\theta}_k | \theta = \tau_k, z_1, \ldots, z_k] \) as the estimation error covariance. The following algorithm can be used to find the MLE iteratively.

\textbf{Algorithm 1 Iterative MLE}

\textbf{Initialization:}
Set \( \hat{\theta}_0 = \mathbb{E}[\theta] \) and \( P_0 = cI_n \) where \( c \) is a constant representing the initial confidence level about how accurate \( \hat{\theta}_0 \) is, i.e., \( c = 0.01 \) for a confident guess or \( c = 100 \) for a very rough guess.

\textbf{Repeat:}

Compute the gain matrix

\[
K_k \leftarrow P_{k-1} (H_k^\top P_{k-1}^{-1}) H_k + \Sigma + (1 - \gamma_k) \Delta_k^{-1}.
\]

Update the estimate and the covariance according to

\[
P_k \leftarrow P_{k-1} - K_k (H_k^\top P_{k-1}^{-1} H_k + \Sigma + (1 - \gamma_k) \Delta_k)^{-1} K_k^\top,
\]

\[
\hat{\theta}_k \leftarrow \hat{\theta}_{k-1} + K_k (x_k - H_k^\top \hat{\theta}_{k-1}) + (1 - \gamma_k) K_k (\tau_k - H_k^\top \hat{\theta}_{k-1} - H_k^{-1} \theta_{k-1}).
\]

\textbf{Until:} \( k = N \), output \( \hat{\theta}_N \) and \( P_N \).

We can have an adaptive estimator in Algorithm 1 with \( \tau_k = H_k^\top \hat{\theta}_{k-1}, \forall k \). Equivalently, the estimator is denoted as

\[
\hat{\theta}_N^\ast := \hat{\theta}_N, \text{ with } \tau_k = H_k^\top \hat{\theta}_{k-1}, \forall k.
\]

The next lemma presents some statistical properties of \( \hat{\theta}_N^\ast \).

\textbf{Theorem 3.5:} Let \( \mathbb{E}[\hat{\theta}_0] = \theta \) and \( P_0 = \infty I_n \). The following statements hold.

1) The adaptive estimator \( \hat{\theta}_N^\ast \) is unbiased, i.e., \( \mathbb{E}[\hat{\theta}_N^\ast] = \theta \).
2) Assume \( H_k (\Sigma + (1 - \gamma_k) \Delta_k)^{-1} H_k^\top \) is invertible, then the error covariance of \( \hat{\theta}_N^\ast \) is

\[
P_N = \left[ \sum_{k=1}^{N} H_k (\Sigma + (1 - \gamma_k) \Delta_k)^{-1} H_k^\top \right]^{-1}.
\]

\textbf{Proof:} 1) Let \( \mathbb{E}[\hat{\theta}_k] = \theta \). From (15), we have

\[
\mathbb{E}[\hat{\theta}_k] = \mathbb{E}[\mathbb{E}[\hat{\theta}_k | \gamma_k]]
\]

\[
= \mathbb{E}[\hat{\theta}_{k-1}] + \Pr\{\gamma_k = 1\} K_k (H_k^\top \theta - \mathbb{E}[\varphi_k] - H_k^\top \mathbb{E}[\hat{\theta}_{k-1}])
\]

\[
+ \Pr\{\gamma_k = 0\} K_k (H_k^\top \mathbb{E}[\hat{\theta}_{k-1}] - H_k^\top \mathbb{E}[\hat{\theta}_{k-1}])
\]

\[
= \theta.
\]
where we denote $K_k = P_{k-1}H_k H_k^T P_{k-1}^{-1}$ and $K_k^d = P_{k-1}^{-1}H_k H_k^T P_{k-1}^{-1} + \Sigma + \Delta^d_k$, since $\mathbb{E}[\theta_0] = \theta$, we can inductively conclude that $\mathbb{E}[\theta_N^\tau] = \theta$.

2) From (13) and (14), we can rewrite the covariance recursion into $P_k^{-1} = P_{k-1}^{-1} + H_k(S + (1 - \gamma_k)\Delta_k)^{-1}H_k^T$. Since $H_k(S + (1 - \gamma_k)\Delta_k)^{-1}H_k^T$ is invertible, the inverse of $P_k$ exists. With the initial condition $P_0 = \alpha I_n$, we can compute the covariance in (17).

We know that the MLE will approach the CRLB as $N \to \infty$ and we are interested in the expression of the CRLB. The following lemma is useful for the derivation of the CRLB.

Lemma 3.6: For any $g_k(\theta)$ defined in (9), it holds that $\mathbb{E}[g_k(\theta)] = 0$.

Proof: By exchanging the summation and the integral, from (9) we have

$$\mathbb{E}[g_k(\theta)] = \int_{\gamma_k=1} H_k \Sigma^{-1} (x_k - H_k^T \theta) f(x_k) dx_k + \int_{\gamma_k=0} H_k (\Sigma + \Delta_k)^{-1} (\tau_k - H_k^T \theta) f(x_k) dx_k$$

$$= \int_{\mathbb{R}^m} H_k \Sigma^{-1} (x_k - H_k^T \theta) f(x_k) dx_k + \int_{\mathbb{R}^m} (H_k (\Sigma + \Delta_k)^{-1} (\tau_k - H_k^T \theta) - H_k \Sigma^{-1} (x_k - H_k^T \theta)) \times f(x_k) \exp(Q_{x_k}(\tau_k, \Delta_k)) dx_k = 0.$$  \hspace{1cm} (18)

The first integral is zero due to $\mathbb{E}[x_k] = H_k^T \theta$. The second integral is also 0 from Lemma 3.1.\hfill\Box

In the next proposition we give the CRLB for any unbiased estimator with the information set $\mathcal{Z}_N$.

Proposition 3.7: The CRLB for any unbiased estimator $\theta_N^b$ based on the information set $\mathcal{Z}_N$ is denoted as $\mathbb{I}_N^b$. Then

$$\text{Var}(\theta_N^b) \geq \mathbb{I}_N^b,$$

where the Fisher information matrix

$$\mathbb{I}_N := \sum_{k=1}^N H_k \Sigma^{-1} H_k^T - \alpha_k H_k \Sigma^{-1} \Omega_k \Sigma^{-1} H_k^T.$$  \hspace{1cm} (20)

Proof: To obtain the CRLB we need to first compute the Fisher information matrix $\mathbb{I}_N := \mathbb{E}[g_N(\theta)^T g_N(\theta)]$ where $g_N(\theta)$ is given in (12). We have

$$\mathbb{E}[g_N(\theta)^T g_N(\theta)] = \mathbb{E} \left[ \left( \sum_{k=1}^N g_k(\theta) \right)^T \left( \sum_{k=1}^N g_k(\theta) \right) \right]$$

$$= \mathbb{E} \left[ \sum_{k=1}^N g_k(\theta) g_k(\theta)^T \right] + \sum_{k \neq j} \mathbb{E}[g_k(\theta)] \mathbb{E}[g_j(\theta)]^T$$

$$= \sum_{k=1}^N \mathbb{E} \left[ g_k(\theta) g_k(\theta)^T \right].$$  \hspace{1cm} (21)

The latter term in (21) results from the fact that $\{g_k(\theta)\}$ is an independent random process. Then from (2), (9) and (22), we have

$$\mathbb{E}[g_N(\theta)^T g_N(\theta)] = \sum_{k=1}^N H_k \Sigma^{-1}$$

$$\times \mathbb{E}[\gamma_k=1] ((x_k - H_k^T \theta)(x_k - H_k^T \theta)^T)(H_k \Sigma^{-1})^T$$

$$+ H_k (\Sigma + \Delta_k)^{-1} \mathbb{E}[\gamma_k=0] ((\tau_k - H_k^T \theta)(\tau_k - H_k^T \theta)^T)$$

$$\times (H_k (\Sigma + \Delta_k)^{-1})^T.$$  \hspace{1cm} (23)

From Lemma 3.1, the first expectation term in (23) is given by

$$\mathbb{E}[\gamma_k=1] ((x_k - H_k^T \theta)(x_k - H_k^T \theta)^T) = \Sigma - \alpha_k [\Omega_k + (\omega_k - H_k^T \theta)(\omega_k - H_k^T \theta)^T].$$  \hspace{1cm} (24)

The second expectation term in (23) is given by

$$\mathbb{E}[\gamma_k=0] ((\tau_k - H_k^T \theta)(\tau_k - H_k^T \theta)^T) = \alpha_k (\tau_k - H_k^T \theta)(\tau_k - H_k^T \theta)^T.$$  \hspace{1cm} (25)

From (23), (24), (25) and Lemma 3.1, we have the CRLB for any unbiased estimator

$$\text{Var}(\theta_N^b) \geq \mathbb{E}[g_N(\theta)^T g_N(\theta)]^{-1} = \mathbb{I}_N^{-1}.$$  \hspace{1cm} \Box

Remark 3.8: The Fisher matrix for the MLE with full measurements is $\sum_{k=1}^N H_k \Sigma^{-1} H_k^T$. In our case that the transmission rate is reduced, the fisher matrix is reduced by $\sum_{k=1}^N \alpha_k H_k \Sigma^{-1} \Omega_k \Sigma^{-1} H_k^T$. This quantitatively reflects the degradation of the estimation performance due to the missing measurements. By designing $\tau_k$ and $\Delta_k$, we can adjust $\alpha_k$ and $\Omega_k$ to meet the estimation quality requirement.

IV. DESIGN OF OPTIMAL PARAMETERS

Generally speaking, the higher the transmission rate is, the lower the estimation covariance of the MLE is. For vector measurements, the mapping between the transmission rate and the error covariance is not one-to-one. We can thus design the parameters $\tau_k$ and $\Delta_k$ in the policy (2) to obtain an optimal tradeoff between the rate and the covariance.

First we quantify the communication cost and the estimation performance and then formulate a constrained optimization problem. As for the communication cost, we treat $\gamma_k = \mathbb{E}[\gamma_k]$ as the transmission intention at time $k$. For the case of the infinite horizon, it is reasonable to uniformly bound the transmission intention such that the communication budget is satisfied. On the other hand, since the MLE reaches the CRLB when $N \to \infty$, we use the CRLB as the estimation performance index. To measure the size of the CRLB, we use the spectrum radius. In this work, we denote the sequence of the parameters as $\tau := \{\tau_k|k = 1, \ldots, N\}, \Delta := \{\Delta_k|k = 1, \ldots, N\}$. Denote the set of all possible sequences $(\tau, \Delta)$ to be $\Xi$. Given a rate constraint $r_0$, we can have a feasible set of sequences $\{(\tau, \Delta)|\exists \gamma_k \leq r_0, \forall k\}$ to satisfy the communication requirement. Thus the question is how to find the optimal solution of $(\tau^*, \Delta^*)$ that minimizes the spectrum radius of the CRLB and satisfies the transmission rate constraint. Denote $W = \lim_{N \to \infty} \frac{1}{N} \mathbb{I}_N$. Mathematically, we are interested in the following optimization problem.
Problem 4.1:

\[
\min_{\Delta \in \Xi} \rho(W^{-1}), \quad (26)
\]
\[
\text{s.t. } \bar{\gamma}_k \leq r_0, \forall k. \quad (27)
\]

Notice that \( \tau \) and \( \Delta \) can be separately designed. We first give a necessary condition for the optimal \( \tau^* \) and then find the optimal \( \Delta \).

Lemma 4.2: The optimal \( \tau^* \) is given by \( \tau_k = H_k^T \theta, \forall k \).

Proof: First we show that \( L(\Delta) := \text{Var}(\hat{\theta}) \) is operator monotone in terms of \( \Delta \). The increasing operator monotonicity is asserted if and only if \( L(\Delta_1) \geq L(\Delta_2) \) for any \( \Delta_1 > \Delta_2 \), where \( \Delta_i := \{ \Delta_k^i \}, i = 1, 2, k = 1, \ldots, N \). Now we prove the monotonicity. We mean the sequence order \( \Delta_1 > \Delta_2 \) by \( \Delta_k^1 - \Delta_k^2 > 0, \forall k \). We have an intermediate result,

\[
\det(I_n + \Sigma(\Delta_k^1)^{-1})^{-\frac{1}{2}} - \det(I_n + \Sigma(\Delta_k^2)^{-1})^{-\frac{1}{2}} > 0, \forall k. \quad (28)
\]

The inequality follows the fact that \( X^{-1} > Y^{-1} > 0 \) and \( \det(Y) > \det(X) > 0 \) if \( Y > X \) where \( X, Y \in S^n_{++} \). Another straightforward result is

\[
\exp(Q_{H_k^T \theta}(\tau_k, \Delta_k^1 + \Sigma)) \geq \exp(Q_{H_k^T \theta}(\tau_k, \Delta_k^2 + \Sigma)), \quad (29)
\]

due to \( v^T v > v^T X v, \forall v \in \mathbb{R}^m \) if \( Y > X \) where \( X, Y \in \mathbb{R}^{m \times m} \). Then from (28) and (29) we have \( L(\Delta_1) > L(\Delta_2) \). We shall prove the lemma by contradiction. If we have the optimal solution \( (\tau^*, \Delta^*) \) to the Problem 4.1 where \( \exists \bar{k}, \tau_k^* \neq H_k^T \theta \), there exists another solution \( (\tau', \Delta') \) where \( \Delta_k' < \Delta_k^* \) and \( \tau_k' = H_k^T \theta \) for the indices \( k \in \{ | \tau_k^* \neq H_k^T \theta \} \) and \( \Delta_k^* \) for the indices \( k \in \{ | \tau_k^* = H_k^T \theta \} \) such that

\[
\det(I_n + \Sigma(\Delta_k')^{-1})^{-\frac{1}{2}} \exp(Q_{H_k^T \theta}(\tau_k', \Delta_k' + \Sigma)) = \det(I_n + \Sigma(\Delta_k^*)^{-1})^{-\frac{1}{2}} \exp(Q_{H_k^T \theta}(\tau_k^*, \Delta_k^* + \Sigma)), \forall k
\]
due to the continuity of \( \det(\cdot) \) and the monotonicity in (28). Thus for both solutions, they lead to the same \( \bar{\gamma}_N \). Due to the operator monotonicity of \( L(\Delta) \), we have \( L(\Delta_k') < L(\Delta_k^*) \) and thus \( \rho(L(\Delta_k')) < \rho(L(\Delta_k^*)) \) which contradicts the optimality assumption of \( (\tau^*, \Delta^*) \). Therefore, the optimal \( \{ \tau_k \} \) must be given by \( \tau_k = H_k^T \theta, \forall k \).

Since practically \( H_k^T \theta \) is not exactly known, we can use \( \bar{\gamma}_k \) instead of \( H_k^T \theta \) to approximate \( H_k^T \theta \). Denote \( W_{+} = W \), where \( \tau_k = H_k^T \theta \). Due to the asymptotic efficiency of the MLE, we equivalently transform Problem 4.1 into

Problem 4.3:

\[
\min_{\Delta} \rho(W_{+}^{-1}), \quad (30)
\]
\[
\text{s.t. } \bar{\gamma}_k \leq r_0, \forall k. \quad (31)
\]

Next we solve the optimization problem of \( \Delta \) via semidefinite programming. First we change the variable \( \Delta_k = \Phi_k \), \( \Phi_k \in \mathbb{R}^{n \times n} \). We have

\[
W_{+}^{-1} \leq t I_n \iff \lambda_{\text{max}}(W_{+}^{-1}) \leq t,
\]

where the \( \lambda_{\text{max}}(W_{+}^{-1}) \) is the maximum eigenvalue of \( W_{+}^{-1} \). Thus (30) and (31) are turned into

Problem 4.4:

\[
\min_{\Delta, t} \quad t
\]
\[
\text{s.t. } W_{+}^{-1} \leq t I_n, \quad (32)
\]
\[
\bar{\gamma}_k \leq r_0, \forall k. \quad (33)
\]

Notice that \( W_{+}^{-1} \leq t I_n \) is equivalent to

\[
\frac{t}{N} I_n - \left( \sum_{k=1}^{N} H_k \Sigma^{-1} H_k^T - M_k \right) \geq 0, \quad (35)
\]

where \( M_k \) is an intermediate matrix variable. It is straightforward to see \( H_k \Sigma^{-1} H_k^T - M_k \geq 0 \) from (36). Then from (35) we have

\[
\left[ \sum_{k=1}^{N} H_k \Sigma^{-1} H_k^T - M_k \right] \quad \text{from } \frac{t}{N} I_n \geq 0
\]

by the Schur complement condition for the positive semidefiniteness. From (34), we can relax (36) into

\[
M_k - (1 - r_0) H_k \Sigma^{-1} (\Phi_k + \Sigma^{-1})^{-1} \Sigma^{-1} H_k^T \geq 0 \quad (37)
\]

Since \( \Phi_k + \Sigma^{-1} > 0 \), by checking Schur complement condition we have

\[
\left[ \frac{1}{1-r_0} (\Phi_k + \Sigma^{-1}) \Sigma^{-1} H_k^T \right] \geq 0, \quad M_k \geq 0.
\]

Next we turn to the constraint in (27). Since \( \bar{\gamma}_k = \alpha_k \) is a log-concave function of \( \Phi_k \), we relax the constraint by replacing \( \bar{\gamma}_k \) with its lower bound. From [14, Lemma 2], we use the following lower bound of \( \bar{\gamma}_k \),

\[
\bar{\gamma}_k \geq 1 - (1 + \text{tr}(\Sigma \Phi_k))^{-\frac{1}{2}}.
\]

Hence we obtain the relaxed constraint

\[
\text{tr}(\Sigma \Phi_k) \leq \left( \frac{2}{1-r_0} \right)^{2} - 1. \quad (38)
\]

Then we have a relaxed SDP optimization problem.

Proposition 4.5: The lower bound of optimal solution \( t \) to Problem 4.4 can be found by solving the following SDP problem.

Problem 4.6:

\[
\min_{t, \{ \Phi_k \}} \quad t
\]
\[
\text{s.t. } \left[ \sum_{k=1}^{N} H_k \Sigma^{-1} H_k^T - M_k \right] \geq 0, \quad (39)
\]
\[
\frac{t}{1-r_0} (\Phi_k + \Sigma^{-1}) \Sigma^{-1} H_k^T \geq 0, \forall k. \quad (40)
\]
\[
\text{tr}(\Sigma \Phi_k) \leq \left( \frac{2}{1-r_0} \right)^{2} - 1, \forall k, \quad (41)
\]
\[
\Phi_k \geq 0, \quad H_k \Sigma^{-1} H_k^T \geq M_k \geq 0, \forall k. \quad (42)
\]

Knowing the optimal solution to Problem 4.6, we have the lower bound of the optimal solution to Problem 4.1. If the optimal \( \Delta^1 \) to the Problem 4.6 satisfies (27), then \( \Delta^* = \Delta^1 \).
Then similarly we can find the upper bound of the optimal solution. From [14, Lemma 2], we use the upper bound of $\gamma_k, \bar{\gamma}_k \leq 1 - \exp(-\frac{1}{2}\text{tr}(\Sigma_k \Phi_k))$.

Proposition 4.7: The upper bound of optimal solution $t$ to Problem 4.4 can be found by solving the following SDP problem.

$$\min \limits_{t, \{\Phi_k\}} t$$

s.t. $\text{tr}(\Sigma_k \Phi_k) \leq -2 \ln(1 - r_0), \forall k,$

(40), (41), (43).

We can solve the problem above to obtain an upper bound of the optimal objective function in Problem 4.1.

V. Numerical Simulations

In this section we present a numerical example to show the main results and illustrate the performance of the proposed scheduling policy. We assume the parameter to be $\theta = [2 \ 1]^\top$. The observation matrix $H_k = [h_1 \ h_2]$ is chosen with each entry uniformly distributed, i.e., $h_1 \sim U[0, 2], h_2 \sim U[1, 3]$. The white Gaussian noise covariance $\Sigma = 1$. In Fig. 2(a) we show the consistency of the MLE estimator in (11) with $\tau_k = 2$ and $\Delta_k = 2$, $\forall k$. We choose the transmission rate constraint $r_0$ to be 0.83. As a reference we also plot the estimate of an LSE with full measurements and the estimate of an MLE with random packet dropouts at the same rate of 0.83. The MLE with random dropouts is equivalent to the LSE with an arbitrary subset of measurements, of which the expected cardinality over the cardinality of the full set is the transmission rate. Moreover, all the estimators are consistent due to the law of large number. In Fig. 2(b) the tradeoff between the transmission rate and the mean squared error (MSE) is shown. The trace of the MSE in dB is defined as $\text{MSE[dB]} = 10 \log_{10}(\text{tr}(\mathbb{E}[\sqrt{N} (\hat{\theta}_N - \theta) (\hat{\theta}_N - \theta)^\top]))$. The time horizon is 500 and $\tau_k = 2$, $\forall k$. The different rates are chosen by adjusting $\Delta_k \in [0.01, 50]$, $\forall k$.

VI. Concluding Remarks

We investigate the optimal parameter estimation under the transmission rate constraint in this work. We design a stochastic scheduling mechanism which defines the importance of each measurement to obtain a better tradeoff between the estimation quality and the transmission rate. By exploiting the resemblance of the proposed stochastic mechanism and the Gaussian density function, we obtain a closed-form data-driven MLE with a subset of measurements. We explicitly give the CRLB of any unbiased estimator with the incomplete measurements. We formulate an optimization problem to search the optimal parameters in the scheduling mechanism to obtain the optimal tradeoff between the estimation performance and the transmission rate requirement.

References


