Optimal Parameter Estimation Under Controlled Communication Over Sensor Networks

Duo Han, Keyou You, Lihua Xie, *Fellow, IEEE*, Junfeng Wu, and Ling Shi

***Abstract*—This paper considers parameter estimation of linear systems under sensor-to-estimator communication constraint. Due to the limited battery power and the traffic congestion over a large sensor network, each sensor is required to reduce the rate of communication between the estimator and itself. We propose an observation-driven sensor scheduling policy such that the sensor transmits only the important measurements to the estimator. Unlike the existing deterministic scheduler, our stochastic scheduling is smartly designed to well compensate for the loss of the Gaussianity of the system. This results in a nice feature that the maximum-likelihood estimator (MLE) is still able to be recursively computed in a closed form, and the resulting estimation performance can be explicitly evaluated. Moreover, an optimization problem is formulated and solved to obtain the best parameters of the scheduling policy under which the estimation performance becomes comparable to the standard MLE with full measurements under a moderate transmission rate. Finally, simulations are included to validate the theoretical results.**

***Index Terms*—Wireless sensor networks, maximum likelihood estimation, Cramer-Rao bounds, event-based communication.**

I. Introduction

DESPITE the benefits such as abundant information without geographical limit that wireless sensor networks provide [1], estimation theory encountered new challenges imposed by wireless communication in recent years. For instance, the measurement quantization effects on the estimation performance have been extensively studied in [2]–[7]. The unreliability of the communication channel such as data packet delay or dropout has also been considered in [8]–[12]. In this paper, we investigate a class of problems called controlled communication for estimation [13]. The concept of controlling communication between the sensors and the remote estimator is rooted in the limited communication resources for each single sensor. What we refer to as communication resources are generally classified into two broad categories, namely, sensor energy and network bandwidth. The most common dilemma the estimation system has to face is that, the sensors are required to temporally and spatially provide as much information as possible to the estimator, but on the other hand the sensors cannot work at full load due to limited battery power or traffic congestion over a network. In other words, provided a limited energy budget, the total number of transmission of one sensor, i.e., the number of data reporting jobs a sensor can do, is finite. Also, when a network of sensors are used, reducing the number of sensors reporting data at the same time relieves the traffic burden and saves the cost of constructing an expensive channel with high bandwidth.

As for the bandwidth-constrained estimation problem, many existing works attack the problem by using one or several bits to represent an analog measurement such that the limited bandwidth is sufficient for a distributed sensor network [2], [6], [7], [14]–[16]. For example, a problem of parameter estimation with binary quantized measurement is studied in [2]. The maximum likelihood estimator (MLE) is conceptually derived and an optimization problem is formulated to find the optimal nonidentical thresholds for binary quantization. The similar idea of binary quantization is applied to the Kalman filtering in [17] and an approximate minimum mean square error (MMSE) estimator is obtained. Marelli *et al.* [7] studied the identification problem for an ARMA model with intermittent and quantized measurements and implemented the MLE estimator using the EM-based algorithm. The energy-constrained estimation problem is also widely studied in [18]–[22]. For example, Li *et al.* [20] explored the tradeoff between the subset of active sensors and the energy used by each active sensor to minimize the estimation error. Mo *et al.* [21] proposed a stochastic sensor selection policy in a sensor network with tree topology to satisfy a limited energy budget.

It is known that the communication unit consumes more power than any other functional block of a sensor [1], [23]. Moreover, a typical 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. Therefore, the burden of limited energy and bandwidth 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.

Manuscript received March 23, 2015; revised July 04, 2015; accepted July 27, 2015. Date of publication August 18, 2015; date of current version November 10, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Mr. Morten Mørup. The work of D. Han and L. Shi was supported by HK Theme-Based Research Grant T23-701/14N. The work of K. You was supported by the National High-Tech R&D Program of China (863 Program) under Grant 2013AA040700. The work of L. Xie was supported in part by the National Research Foundation of Singapore under Grants NRF2011NRF-CRP001-090 and NRF2013EWT-EIRP004-012, and in part by the Natural Science Foundation of China under NSFC 61120106011. The work of J. Wu was supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

D. Han and L. Shi are with the Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: dhanaa@ust.hk; eesling@ust.hk).

K. You is with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: youky@tsinghua.edu.cn).

L. Xie is with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798 Singapore (e-mail: elhxie@ntu.edu.sg).

J. Wu is with the Department of Automatic Control, The Royal Institute of Technology (KTH), Stockholm SE-100 44, Sweden (e-mail: junfengw@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2469639

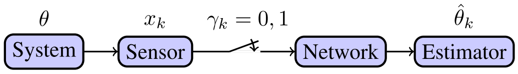


Fig. 1. Parameter estimation framework.

We focus on the parameter estimation under the constrained transmission rate in this paper. The sensor measurements are used to estimate an unknown but deterministic vector parameter. However, not all the measurements are transmitted to the estimator for some reasons, such as extending a sensor's life span, relieving the traffic congestion, etc. An alternative countermeasure is that the sensor is allowed to send its measurements periodically. The length of the period is designed based on the desired tradeoff between the performance quality and the communication cost. In this method, the sensor treats all the measurements with *equal* importance. However, this method is shown to provide a worse tradeoff than an adaptive one which can *differentiate* the importance of each sensor [24], [25]. A better idea is that by defining some metric to measure the importance, the sensor at each time intentionally discards the unimportant data packets to save their communication resources. This concept has been widely used for estimating the state of a dynamical system [25]–[28]. A closely related work [29] 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. Their MLE lacks a simple form and is computationally demanding.

We consider the remote estimation scheme shown in Fig. 1, where the sensor observes the vector parameter through a time-invariant or time-varying observation matrix with additive Gaussian noise. To save the communication cost, the sensor decides whether to send the measurement x_k to the estimator. In the spirit of [27] which dealt with the MMSE state estimation of a linear time-invariant system, we propose a stochastic scheduling solution to the transmission rate constrained parameter estimation problem. The solution results in a closed-form MLE and a recursive algorithm to compute the MLE, which possesses a main advantage over the deterministic threshold-type mechanism in [29]. Unlike [27], we analytically present extensive asymptotic results on the proposed MLE which show the consistency and asymptotic normality. We also formulate an optimization problem to find the optimal parameters in the scheduling policy. The main contributions of this work are summarized as follows.

1) We design a stochastic scheduling policy to reduce the transmission rate while preserving sufficient estimation accuracy. Then, we derive a *closed-form* MLE which is impossible in the existing work and present a recursive MLE algorithm.
2) Though it is biased in general, the MLE is proved to be asymptotically unbiased and consistent. We derive an analytical expression of the Cramér-Rao lower bound (CRLB), which the covariance of the MLE asymptotically reaches under some mild assumption. Moreover, we extend the result into the case where the observation matrix is random.
3) The proposed stochastic scheduling policy with two new design parameters can be adjusted to achieve an *arbitrary* tradeoff between the estimation quality and the communication. A general optimization problem is formulated, with respect to the new design parameters, to find the optimal tradeoff balance. The optimization problem is successfully relaxed and the tight upper and lower bounds of the optimal solution are solved using semidefinite programming (SDP).
4) When the prior knowledge of the parameter to be estimated is available to the estimator, we study the maximum *a posteriori* estimation problem and more generally, the Bayes estimation problem.

The remainder of this paper is organized as follows. In Section II we describe the parameter estimation problem of interest and present a novel scheduling solution. Then we solve the MLE in a closed-form and study some variations in Section III. In Section IV we carry out asymptotic analysis of the derived estimator. In the following section we discuss how to search the optimal parameters in the scheduling policy by formulating an optimization problem. In Section VI we discuss how to incorporate prior knowledge of the parameter to be estimated into the MLE. Some numerical results and concluding remarks are given in Section VII and VIII.

Notations: Let $x \in \mathbb{R}^n$ be a random vector on a probability space $(\Omega, \mathcal{F}, \Pr)$. We use $f_x(\cdot)$ to represent a probability density function (abbreviated as pdf) of x. $x_n \xrightarrow{a.s.} x$ means that a sequence of random vectors indexed by n converges almost surely to the random vector x. $x_n \xrightarrow{d} x$ means that a sequence of random vectors indexed by n converges in distribution to the random vector x. $\nabla(g)$ and $\nabla^2(g)$ are the gradient and the Hessian matrix of a scalar function g. $\mathbf{I}_n$ is the identity matrix of rank n. $\det(A)$ is the determinant of the square matrix A. $\rho(A)$ is the spectral radius of the matrix A, i.e., the largest eigenvalue in magnitude of A. $\mathbb{R}_+^{n\times n}$ and $\mathbb{R}_{++}^{n\times n}$ are the set of $n \times n$ positive semidefinite and definite matrices. To compare two $n \times n$ matrices, we denote $A \geq B$ if $A - B \in \mathbb{R}_+^{n\times n}$, $A > B$ likewise. $Y^\dagger$ is the Moore-Penrose pseudoinverse of $Y \in \mathbb{R}^{n\times m}$. $\|X\|_\infty$ is the ∞-norm of a matrix X, i.e., the maximum of the absolute row sums. Given a positive definite matrix X, $\|y\|_X := (y^\top X y)^{\frac{1}{2}}$ is the weighted norm of the vector y.

II. Problem Statement

Consider a parameter estimation problem of estimating $\theta \in \mathbb{R}^n$ described as

$$x_k = H_k^\top \theta + \nu_k, 1 \leq k \leq N \tag{1}$$

where $x_k \in \mathbb{R}^m$ is the measurement and $\nu_k \in \mathbb{R}^m$ is the 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 $\sup_k \|H_k\|_\infty < \infty$. The parameter θ is unknown and to be identified. The sensor measures the contaminated value x_k of the parameter θ and sends the measurement to a remote estimator for estimating θ.

Taking both the estimation quality and the communication cost into account, we choose a subset of measurements to transmit for an estimation task, according to a certain criterion. We consider designing a sensor scheduling policy such that those more *important* measurements are more *likely* to be received by the estimator. Consequently, we expand the sensor lifetime in an average sense. Let γ_k be the indicator whether the sensor is authorized to transmit data at time k, i.e., permitted when $\gamma_k = 1$ and denied when $\gamma_k = 0$. Many criteria for determining which measurement is important have been proposed. For example, a sensor scheduling policy $\{\gamma_k\}$ is considered in [29]:

$$\gamma_k = \begin{cases} 0, & \text{if } |\frac{x_k - \tau_k}{\sigma}| \leq \delta, \\ 1, & \text{otherwise,} \end{cases} \tag{2}$$

where τ_k and σ are given. When the real function of the scalar measurement x_k exceeds the threshold δ, a transmission is triggered.

In this work we consider the following stochastic sensor scheduling policy $\{\gamma_k\}$: given a specific x_k, the transmission indicator γ_k is a Bernoulli random variable with the probability of failure $\exp(-(1/2)\|x_k - \tau_k\|^2_{\Delta_k^{-1}})$, i.e.,

$$\gamma_k|_{x_k=x} \sim Bern\left(1 - \exp\left(-\frac{1}{2}\left\|x - \tau_k\right\|^2_{\Delta_k^{-1}}\right)\right), \tag{3}$$

where the sequence of vectors $\{\tau_k \in \mathbb{R}^m | 1 \leq k \leq N\}$ and the sequence of positive definite matrices $\{\Delta_k \in \mathbb{R}^{m\times m}_{++} | 1 \leq k \leq N\}$ are the parameters to be designed. Note that the conditional random variable $\gamma_k|_{x_k=x}$ is defined with a known distribution but the probability of success of γ_k itself is unknown, which will be shown later.

An observation from (3) is that the probability of $\gamma_k = 1$ is lower when x_k is closer to τ_k, namely, a transmission is less likely when x_k is closer to τ_k. The intuition behind (3) is that, from a viewpoint of the estimator, $\gamma_k = 0$ implies that x_k is around the known reference τ_k with a high probability, and thus the estimator is still able to make a good guess of x_k even when the measurement is dropped for communication saving. The parameters τ_k and Δ_k, of which the estimator has a knowledge, play roles of *known reference* and *scaling factor*, respectively. It is worthwhile to notice that the probability of $\gamma_k = 1$ is a binary function of the value of the predefined function of x_k in (2) while the probability of $\gamma_k = 1$ is a continuous function of the value of the predefined function of x_k in (3). The exponential function of τ_k and Δ_k in (3) resembling the Gaussian pdf renders the derivation of the MLE tractable, which we shall see later.

Remark 1: The drawback of the deterministic scheduling policy in [29] is the lack of an explicit expression of the resulting MLE. In the case of vector measurements, the computational complexity of the multivariate integration is demanding. This motivates us to find a stochastic policy to reduce the computation burden and also achieve a good tradeoff between estimation and communication.

The operation principle of the sensor is described as follows. The sensor collects the observation x_k and computes the real-valued function $\eta = 1 - \exp(-(1/2)\|x_k - \tau_k\|^2_{\Delta_k^{-1}})$. Then the sensor generates a Bernoulli random variable γ_k with the probability of success η to determine whether to transmit x_k.

To facilitate the derivation, we denote the information set at the estimator side as

$$\mathcal{Z}_N := \{z_1, \ldots, z_k, \ldots, z_N\}, \tag{4}$$

where

$$z_k := \begin{cases} x_k, & \text{when } \gamma_k = 1, \\ 0, & \text{when } \gamma_k = 0. \end{cases}$$

We use $z_k = 0$ to represent the event of $\gamma_k = 0$. Notice that there is a notation conflict between the case where $x_k = 0$ and $\gamma_k = 1$, and the case where $\gamma_k = 0$. Since the probability measure of $x_k = 0$ when $\gamma_k = 1$ is 0, the fact that $z_k = 0$ only represents the event of $\gamma_k = 0$ will not affect the analytical results in the sequel. Furthermore, the average transmission rate is denoted to be

$$\bar{\gamma}_N = \frac{1}{N}\sum_{k=1}^{N}\gamma_k. \tag{5}$$

The remaining questions after we propose the scheduling policy for the parameter estimation problem are

1) Under the stochastic scheduling policy, how shall we calculate the MLE based on the information set $\mathcal{Z}_N$? And how shall we evaluate the average transmission rate $\bar{\gamma}_N$?
2) With respect to an increasing number of measurements, what is the asymptotic stochastic properties of the parameter estimation, in terms of the unbiasedness, consistency and normality?
3) Is it possible to jointly optimize the estimation quality and the transmission rate by making use of the design freedom τ_k and Δ_k? If so, how do we solve the resulting optimization problem?
4) If there is some prior statistical knowledge of the unknown parameter, how can we incorporate that information into the parameter estimation process?

In the following sections, we will give answers to all these questions.

III. Maximum Likelihood Estimation With Scheduled Measurements

In this section we present the MLE with the information set in (4). 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.,

$$f_{x_k}(x) := \mathcal{N}\left(H_k^\top\theta, \Sigma\right) = \frac{\exp\left(-\frac{1}{2}\left\|x_k - H_k^\top\theta\right\|^2_{\Sigma^{-1}}\right)}{\sqrt{(2\pi)^n \det(\Sigma)}}. \tag{6}$$

Let us define an auxiliary function for the pdf $f_{x_k}(x)$ given a pair of (τ_k, Δ_k),

$$d_{(\tau_k,\Delta_k)}(x) := \frac{1}{\alpha_k}\exp\left(-\frac{1}{2}\|x - \tau_k\|^2_{\Delta_k^{-1}}\right)$$

where

$$\alpha_k = \det(\mathbf{I}_n + \Sigma\Delta_k^{-1})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left\|H_k^\top\theta - \tau_k\right\|^2_{\Delta_k+\Sigma^{-1}}\right). \tag{7}$$

We need the following intermediate result to facilitate the derivation of the MLE, which shows that the product of the auxiliary function and the density function $f_{x_k}(x)$ is still Gaussian.

Lemma 1: For any given pair of (τ_k, Δ_k), the following equality holds:

$$d_{(\tau_k,\Delta_k)}(x) \cdot f_{x_k}(x) = \mathcal{N}(\omega_k, \Omega_k), \tag{8}$$

where

$$\omega_k = H_k^\top\theta + \Sigma(\Sigma+\Delta_k)^{-1}(\tau_k - H_k^\top\theta), \tag{9}$$
$$\Omega_k = \Sigma - \Sigma(\Sigma+\Delta_k)^{-1}. \tag{10}$$

Proof: The result is proved by completion of squares,

$$\begin{aligned} &d_{(\tau_k,\Delta_k)}(x) \cdot f_{x_k}(x) \\ &= \frac{1}{\alpha_k} \cdot \frac{\exp\left(-\frac{1}{2}\|x-\tau_k\|^2_{\Delta_k^{-1}} - \frac{1}{2}\left\|x - H_k^\top\theta\right\|^2_{\Sigma^{-1}}\right)}{\sqrt{(2\pi)^n \det(\Sigma)}} \\ &= \frac{\exp\left(-\frac{1}{2}\|x-\omega_k\|^2_{\Omega_k^{-1}}\right)}{\sqrt{(2\pi)^n\det(\Omega_k)}}. \end{aligned}$$

■It is shown that the product of two Gaussian-like function is still a Gaussian-like function. The constant α_k is a normalized constant to ensure $f_{\tilde{x}_k}(x)$ to be a Gaussian pdf. The transformed mean ω_k and covariance Ω_k play important roles in the subsequent sections. Now we have an immediate result on the probability of a measurement being selected to transmitted.

Lemma 2: Given the stochastic sensor scheduling rule in (3), the probability of transmission permission for the sensor at time k is given by

$$\Pr\{\gamma_k = 1\} = 1 - \det\left(\mathbf{I}_n + \Sigma\Delta_k^{-1}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left\|H_k^\top\theta - \tau_k\right\|^2_{\Delta_k+\Sigma^{-1}}\right). \tag{11}$$

Proof: The probability of success of the Bernoulli random variable γ_k depends on the Gaussian random variable x_k from (3). To obtain the probability of $\gamma_k = 1$, we take the expectation of the probability of success of the conditional random variable $\gamma_k|_{x_k=x}$ in (3) over x_k. Thus we have that

$$\begin{aligned} \Pr\{\gamma_k = 1\} &= \int_{-\infty}^{+\infty} \Pr(\gamma_k = 1|x_k = x) f_{x_k}(x)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \left[1 - \exp\left(-\frac{1}{2}\|x-\tau_k\|^2_{\Delta_k^{-1}}\right)\right] f_{x_k}(x)\mathrm{d}x \\ &= 1 - \alpha_k \int_{\mathbb{R}^m} f_{\tilde{x}_k}(x)\mathrm{d}x. \end{aligned}$$

The first equality is due to Bayes' theorem and the third equality is due to $\int f_{\tilde{x}_k}(x)\mathrm{d}x = 1$. ■

By a little abuse of notation, we use $f_{x_k}(x)$ and $f(x_k)$ interchangeably to denote the pdf of x_k. From Lemma 2, the joint pdf of $\mathcal{Z}_N$ with the unknown parameter θ is given by

$$f_{\mathcal{Z}_N}(z_1,\ldots,z_N;\theta) = \prod_{k=1}^N f_{x_k}(x)^{\gamma_k} \Pr\{\gamma_k = 0\}^{1-\gamma_k}. \tag{12}$$

Then the log-likelihood function is written as

$$l_{\mathcal{Z}_N}(z_1,\ldots,z_N;\theta) = \log f_{\mathcal{Z}_N}(z_1,\ldots,z_N;\theta).$$

Abbreviating $l_{\mathcal{Z}_N}(z_1,\ldots,z_N;\theta)$ to $l_N(\theta)$ for simplicity, then we have,

$$\begin{aligned} l_N(\theta) = \sum_{k=1}^N \Big[&\gamma_k \log\left((2\pi)^n \det(\Sigma)\right)^{-\frac{1}{2}} \\ &- \frac{1}{2}\gamma_k\left\|H_k^\top\theta - x_k\right\|^2_{\Sigma^{-1}} \\ &+ (1-\gamma_k)\log\left(\det\left(\mathbf{I}_n + \Sigma\Delta_k^{-1}\right)^{-\frac{1}{2}}\right) \\ &- \frac{1}{2}(1-\gamma_k)\left\|H_k^\top\theta - \tau_k\right\|^2_{\Delta_k+\Sigma^{-1}}\Big]. \end{aligned} \tag{13}$$

For the ease of derivation, we define the following functions of θ,

$$\begin{aligned} g_k(\theta) :=\ &\gamma_k H_k\Sigma^{-1}\left(x_k - H_k^\top\theta\right) \\ &+ (1-\gamma_k)H_k(\Sigma+\Delta_k)^{-1}\left(\tau_k - H_k^\top\theta\right), \end{aligned} \tag{14}$$
$$h_k(\theta) := \gamma_k H_k\Sigma^{-1}H_k^\top + (1-\gamma_k)H_k(\Sigma+\Delta_k)^{-1}H_k^\top. \tag{15}$$

We are now ready to present the MLE under the information set (4). We shall see that the resemblance between the threshold in the selection policy (3) and the Gaussian distribution facilitates the MLE derivation.

Theorem 1: Consider the estimation problem in (1) and the scheduling policy in (3). The MLE $\hat{\theta}_N$ based on $\mathcal{Z}_N$ is given as

$$\begin{aligned} \hat{\theta}_N = &\left[\sum_{k=1}^N \gamma_k H_k\Sigma^{-1}H_k^\top + (1-\gamma_k)H_k(\Sigma+\Delta_k)^{-1}H_k^\top\right]^\dagger \\ &\times \left[\sum_{k=1}^N \gamma_k H_k\Sigma^{-1}x_k + (1-\gamma_k)H_k(\Sigma+\Delta_k)^{-1}\tau_k\right]. \end{aligned} \tag{16}$$

Proof: Taking the gradient of $l_N(\theta)$ in (13) over θ, we have

$$\nabla\left(l_N(\theta)\right) = \sum_{k=1}^N g_k(\theta). \tag{17}$$

Also taking the Hessian matrix of $l_{\mathcal{Z}_N}(z_1,\ldots,z_N;\theta)$, we have

$$\nabla^2\left(l_N(\theta)\right) = -\sum_{k=1}^N h_k(\theta). \tag{18}$$

Note that $\nabla^2(l_N(\theta)) < 0$ since $\Sigma, \Delta_k > 0$. Thus $l_{\mathcal{Z}_N}(z_1,\ldots,z_N;\theta)$ is a concave function over $\theta \in \mathbb{R}^n$. Letting $\nabla(l_N(\theta)) = 0$, we obtain $\hat{\theta}_N = \arg\max_\theta l_N(\theta)$ is the MLE which completes the proof. ■

Remark 2: In [29] the MLE with the deterministic scheduler cannot be explicitly written and some algorithms are proposed to numerically find the MLE. The high computational complexity 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) [30], we can write the MLE in a recursive way. Denote $\hat{\theta}_k := \mathbb{E}[\theta|z_1,\dots,z_k]$ as the estimate of θ after k measurements and $P_k := \mathbb{E}[(\theta-\hat{\theta}_k)(\theta-\hat{\theta}_k)^\top|z_1,\dots,z_k]$ as the estimation error covariance. The following algorithm can be used to find the MLE iteratively.[1]

Algorithm 1: Iterative MLE

Initialization:

Initialize $\hat{\theta}_0 = 0$ and $P_0 = c\mathrm{I}_n, c > 0$.

Repeat:

Compute the gain matrix

$$K_k \leftarrow P_{k-1}H_k\left(H_k^\top P_{k-1}H_k + \Sigma + (1-\gamma_k)\Delta_k\right)^{-1}. \quad (19)$$

Update the estimate and the covariance according to

$$P_k \leftarrow P_{k-1} - K_k\left(H_k^\top P_{k-1}H_k + \Sigma + (1-\gamma_k)\Delta_k\right)K_k^\top, \quad (20)$$

$$\hat{\theta}_k \leftarrow \hat{\theta}_{k-1} + \gamma_k K_k\left(x_k - H_k^\top\hat{\theta}_{k-1}\right) + (1-\gamma_k)K_k\left(\tau_k - H_k^\top\hat{\theta}_{k-1}\right)\Big]. \quad (21)$$

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

It is worth noting that the MLE is biased and $\mathbb{E}[\hat{\theta}_N]$ depends on τ_k. To see it clearly, we first show that $\mathbb{E}[g_k(\theta)] = 0$, where $g_k(\theta)$ is defined in (14). The proof is reported in the Appendix.

Lemma 3: Consider $g_k(\theta)$ defined in (14). For any $1 \le k \le N$, $\mathbb{E}[g_k(\theta)] = 0$.

Now, we use a simple example to illustrate that the estimator $\hat{\theta}_N$ is generically biased. Considering that the MLE with only one observation $\{z_1\}$ as $\hat{\theta}_1$, from (16) we have

$$\hat{\theta}_1 = \left[\gamma_1 H_1\Sigma^{-1}H_1^\top + (1-\gamma_1)H_1(\Sigma+\Delta_1)^{-1}H_1^\top\right]^\dagger \times\left[\gamma_1 H_1\Sigma^{-1}x_1 + (1-\gamma_1)H_1(\Sigma+\Delta_1)^{-1}\tau_1\right]. \quad (22)$$

Then we have

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_1] &= \int_{\{\gamma_1=1\}} \left(H_1\Sigma^{-1}H_1^\top\right)^\dagger H_1\Sigma^{-1}\left(x_1 - H_1^\top\theta + H_1^\top\theta\right)\mathrm{d}x_1 \\
&\quad + \int_{\{\gamma_1=0\}} \left(H_1(\Sigma+\Delta_1)^{-1}H_1^\top\right)^\dagger H_1(\Sigma+\Delta_1)^{-1} \\
&\quad \times\left(\tau_1 - H_1^\top\theta + H_1^\top\theta\right)\mathrm{d}x_1 \\
&= \theta + \left(H_1\Sigma^{-1}H_1^\top\right)^\dagger\mathbb{E}[g_1(\theta)] \\
&\quad + \int_{\{\gamma_1=0\}} \left[\left(H_1(\Sigma+\Delta_1)^{-1}H_1^\top\right)^\dagger - \left(H_1\Sigma^{-1}H_1^\top\right)^\dagger\right] \\
&\quad \times\left(\tau_1 - H_1^\top\theta\right)\mathrm{d}x_1 \\
&= \theta + \alpha_1\Lambda_1\left(\tau_1 - H_1^\top\theta\right),
\end{aligned} \quad (23)$$

where $\Lambda_1 := \left(H_1(\Sigma+\Delta_1)^{-1}H_1^\top\right)^\dagger - \left(H_1\Sigma^{-1}H_1^\top\right)^\dagger$. The second equality is due to Lemma 1 and 3. Note that even the one-measurement estimator is biased unless $\tau_1 = H_1^\top\theta$. In fact, we do not know the exact value of θ.

To maintain the unbiasedness of the MLE under the proposed policy in (3), we let $\tau_k = H_k^\top\hat{\theta}_{k-1}, \forall k$ for (21) in Algorithm 1. To be specific, the adaptive estimator is denoted as

$$\hat{\theta}_N^* := \hat{\theta}_N, \text{ with } \tau_k = H_k^\top\hat{\theta}_{k-1}, \forall k. \quad (24)$$

The next theorem presents some statistical properties of $\hat{\theta}_N^*$.

Theorem 2: Let $\mathbb{E}[\hat{\theta}_0] = \theta$ and assume P_0 to be invertible. The following statements hold.

1) The adaptive estimator $\hat{\theta}_N^*$ is unbiased, i.e., $\mathbb{E}[\hat{\theta}_N^*] = \theta$.
2) The error covariance of $\hat{\theta}_N^*$ is

$$P_N = \left[P_0^{-1} + \sum_{k=1}^{N} H_k\left(\Sigma + (1-\gamma_k)\Delta_k\right)^{-1}H_k^T\right]^{-1}. \quad (25)$$

Proof:

1) Let $\mathbb{E}[\hat{\theta}_{k-1}] = \theta$. From (21), we have

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_k] &= \mathbb{E}\left[\mathbb{E}[\hat{\theta}_k|\gamma_k]\right] \\
&= \mathbb{E}[\hat{\theta}_{k-1}] + \Pr\{\gamma_k = 1\}K_k^1 \\
&\quad \times\left(H_k^\top\theta - \mathbb{E}[\nu_k] - H_k^\top\mathbb{E}[\hat{\theta}_{k-1}]\right) \\
&\quad + \Pr\{\gamma_k = 0\}K_k^0\left(H_k^\top\mathbb{E}[\hat{\theta}_{k-1}] - H_k^\top\mathbb{E}[\hat{\theta}_{k-1}]\right) \\
&= \theta,
\end{aligned}$$

where we denote $K_k^1 = P_{k-1}H_k(H_k^\top P_{k-1}H_k + \Sigma)^{-1}$ and $K_k^0 = P_{k-1}H_k(H_k^\top P_{k-1}H_k + \Sigma + \Delta_k)^{-1}$. Since $\mathbb{E}[\hat{\theta}_0] = \theta$, we inductively conclude that $\mathbb{E}[\hat{\theta}_N^*] = \theta$.

2) From (19) and (20), we can rewrite the covariance recursion as

$$P_k^{-1} = P_{k-1}^{-1} + H_k\left(\Sigma + (1-\gamma_k)\Delta_k\right)^{-1}H_k^\top.$$

Assuming the initial covariance P_0 to be invertible, we can compute the covariance in (25). ■

IV. Asymptotic Analysis of the MLE With Scheduled Measurements

In this section we study the asymptotic properties of the MLE in (16). We show that $\hat{\theta}_N$ is *consistent* and *asymptotically normal* with an *explicit* form of the stationary covariance. Moreover, we manage to calculate the average transmission rate. First we need to introduce the following persistent excitation [31] assumption which is essential for the asymptotic convergence of the MLE with full measurements.

Assumption 1: There exists a real number $\varsigma > 0$ such that

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} H_k\Sigma^{-1}H_k^\top \ge \varsigma\mathrm{I}_n. \quad (26)$$

In the next proposition we give the CRLB for any unbiased estimator with the information set $\mathcal{Z}_N$. The proof is reported in the Appendix.

Proposition 1: Let the Fisher information matrix to be

$$\mathfrak{I}_N = \sum_{k=1}^{N} H_k\Sigma^{-1}H_k^\top - \alpha_k H_k\Sigma^{-1}\Omega_k\Sigma^{-1}H_k^\top. \quad (27)$$

[1]In the initialization 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.

Then the CRLB for any unbiased estimator θ_N^u based on $\mathcal{Z}_N$ in (4) is given by $\Im_N^{-1}$, i.e.,

$$\mathrm{Var}\,(\theta_N^u) \geq \Im_N^{-1}, \tag{28}$$

Next we present the asymptotic properties of the MLE. We show that the covariance of the MLE asymptotically reaches the CRLB and thus the MLE is asymptotically optimal from Proposition 1. The mild condition of persistent excitation in Assumption 1 is needed to guarantee the consistency and asymptotic normality.

Theorem 3: Under Assumption 1 and the assumptions of $\sup_k \|H_k\| \leq \infty$ and $\sup_k \|\tau_k\| \leq \infty$, the MLE $\hat{\theta}_N$ in (16) has the following asymptotic statistical properties.

1) Consistency: $\hat{\theta}_N \xrightarrow{a.s.} \theta$.
2) Asymptotic unbiasedness: $\mathbb{E}[\hat{\theta}_N] \xrightarrow{a.s.} \theta$.
3) Asymptotic normality: assume that W exists, $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, W^{-1})$, where

$$W = \lim_{N\to\infty} \frac{1}{N}\Im_N. \tag{29}$$

4) Average transmission rate: assume that $\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N \alpha_k$ exists, the average transmission rate is

$$\bar{\gamma}_N \xrightarrow{a.s.} 1 - \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N \alpha_k. \tag{30}$$

The proof is reported in the Appendix.

Remark 3: The Fisher matrix for the MLE with full measurements is $\sum_{k=1}^N H_k\Sigma^{-1}H_k^\top$. Under the scheduling policy (3), the fisher matrix is reduced by $\sum_{k=1}^N \alpha_k H_k\Sigma^{-1}\Omega_k\Sigma^{-1}H_k^\top$. This *quantitatively* reflects the degradation of the estimation performance due to the missing measurements. By designing τ_k and Δ_k, we can adjust α_k and Ω_k to meet the estimation quality requirement.

When the MLE is in the recursive form in Algorithm 1, we show the asymptotic properties of the recursive MLE $\hat{\theta}_N^*$ in (24).

Theorem 4: Under Assumption 1 and the assumptions of $\sup_k \|H_k\| \leq \infty$ and $\sup_k \|\tau_k\| \leq \infty$, let $\tau_k = H_k^\top\hat{\theta}_{k-1}$ and the MLE $\hat{\theta}_N^*$ in (16) has the following asymptotic statistical properties.

1) Consistency: $\hat{\theta}_N^* \xrightarrow{a.s.} \theta$.
2) Asymptotic normality: assume that W_+ exists, $\sqrt{N}(\hat{\theta}_N^* - \theta) \xrightarrow{d} \mathcal{N}(0, W_+^{-1})$, where

$$W_+ = \lim_{N\to+\infty}\frac{1}{N}\sum_{k=1}^N H_k\Sigma^{-1}H_k^\top - \det\left(\mathbf{I}_n + \Sigma\Delta_k^{-1}\right)^{-\frac{1}{2}} H_k\Sigma^{-1}\Omega_k\Sigma^{-1}H_k^\top \tag{31}$$

3) Average transmission rate: assume that $\lim_{N\to+\infty}\frac{1}{N}\sum_{k=1}^N \det\left(\mathbf{I}_n + \Sigma\Delta_k^{-1}\right)^{-\frac{1}{2}}$ exists,

$$\bar{\gamma}_N^* \xrightarrow{a.s.} 1 - \lim_{N\to+\infty}\frac{1}{N}\sum_{k=1}^N \det\left(\mathbf{I}_n + \Sigma\Delta_k^{-1}\right)^{-\frac{1}{2}}. \tag{32}$$

The proof is reported in the Appendix.

Remark 4: Comparing Theorem 3 and 4, we see that $\bar{\gamma}_N^* \leq \bar{\gamma}_N$ and $W_+^{-1} \geq W^{-1}$ for the same Δ_k. Notice that the low transmission rate and the small error covariance are favorable. Though not clear now, we will show that $\tau_k = H_k^\top\hat{\theta}_{k-1}$ is generally asymptotically optimal in terms of the rate and the error covariance later.

In the aforementioned asymptotic results, $\{H_k\}$ is assumed to be a deterministic sequence. The asymptotic properties still apply if we give a sufficient condition of wide-sense stationary ergodicity for $\{H_k\}$. The following result can be easily proved by using the Birkhoff's ergodic theorem.

Corollary 1: If $\{H_k\}$ is a wide-sense stationary ergodic random process with uniformly bounded ∞-th moment, then the asymptotic normality in Theorem 3 and Theorem 4 still holds.

V. Design of Optimal Parameters in the Scheduling Policy

For vector measurements, the mapping between the transmission rate and the error covariance is *not* one-to-one. We can thus design the parameters τ_k and Δ_k in the policy (3) to obtain an optimal tradeoff between the rate and the covariance. We quantify the communication cost and the estimation performance and then formulate a constrained optimization problem. As for the communication cost, we treat $\bar{\gamma}_k = \mathbb{E}[\gamma_k]$ 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 spectral radius. In this work, we denote the sequence of the parameters as

$$\tau := \{\tau_1, \ldots, \tau_N\},$$
$$\Delta := \{\Delta_1, \ldots, \Delta_N\}.$$

Denote the set of all possible sequences (τ, Δ) to be Ξ. Given a rate constraint r_0, we can have a feasible set of sequences

$$\{(\tau, \Delta) \in \Xi | \bar{\gamma}_k \leq r_0, \forall k\}$$

to satisfy the communication requirement. Thus the question is how to find the optimal solution of (τ^*, Δ^*) that minimizes the spectral radius of the CRLB and satisfies the transmission rate constraint. Mathematically, we are interested in the following optimization problem.

Problem 1:

$$\min_{(\tau,\Delta)\in\Xi} \rho(W^{-1}), \tag{33}$$

$$\text{s.t. } \bar{\gamma}_k \leq r_0, \forall k, \tag{34}$$

where $\rho(W^{-1})$ is the spectral radius of W^{-1}. We first give a necessary condition for the optimal τ^* and then find the optimal Δ, which enables that τ and Δ can be separately designed.

Lemma 4: The optimal τ^* is given by $\tau_k = H_k^\top\theta, \forall k$. The proof is given in the Appendix. Letting $\tau_k = H_k^\top\theta$, we transform Problem 1 into

Problem 2:

$$\min_{\Delta} \rho\left(W_+^{-1}\right), \tag{35}$$

$$\text{s.t. } \bar{\gamma}_k \leq r_0, \forall k. \tag{36}$$

Practically, $H_k^\top \theta$ is not exactly known. Since we are more concerned about the performance after a long period, we use $\tau_k = H_k\hat{\theta}_{k-1}$ to approximate $H_k^\top \theta$ due to the consistency of $\hat{\theta}_N$ given by Theorem 4. By introducing a random variable t, we can rewrite Problem 2 into

Problem 3:

$$\min_{\Delta,t}\ t \tag{37}$$

$$\text{s.t. } W_+^{-1} \leq t\mathrm{I}_n, \tag{38}$$

$$\bar{\gamma}_k \leq r_0, \forall k. \tag{39}$$

By changing the variable $\Delta_k = \Phi_k^{-1}, \Phi_k \in \mathbb{R}_{++}^{n\times n}$, we can see that $W_+^{-1} \leq t\mathrm{I}_n$ is equivalent to

$$\frac{t}{N}\mathrm{I}_n - \left(\sum_{k=1}^{N} H_k\Sigma^{-1}H_k^\top - M_k\right)^{-1} \geq 0, \tag{40}$$

$$H_k\Sigma^{-1}H_k^\top \geq M_k \geq \alpha_k H_k\Sigma^{-1}\left(\Phi_k+\Sigma^{-1}\right)^{-1}\Sigma^{-1}H_k^\top, \forall k \tag{41}$$

where M_k is an intermediate matrix variable. It is straightforward to see $H_k\Sigma^{-1}H_k^\top - M_k \geq 0$ from (41). Then from (40) we have

$$\begin{bmatrix} \sum_{k=1}^{N} H_k\Sigma^{-1}H_k^\top - M_k & \mathrm{I}_n \\ \mathrm{I}_n & \frac{t}{N}\mathrm{I}_n \end{bmatrix} \geq 0$$

by the Schur complement condition for the positive semidefiniteness. From (39), we can transform (41) into

$$M_k - (1-r_0)H_k\Sigma^{-1}\left(\Phi_k+\Sigma^{-1}\right)^{-1}\Sigma^{-1}H_k^\top \geq 0. \tag{42}$$

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

$$\begin{bmatrix} \frac{1}{1-r_0}\left(\Phi_k+\Sigma^{-1}\right) & \Sigma^{-1}H_k^\top \\ H_k\Sigma^{-1} & M_k \end{bmatrix} \geq 0, M_k \geq 0.$$

Next we turn to the constraint in (34). Since $\bar{\gamma}_k = \alpha_k$ is a log-concave function of Φ_k, we relax the constraint by replacing $\bar{\gamma}_k$ with its lower bound. From [27, Lemma 2], we use the following lower bound of $\bar{\gamma}_k$,

$$\bar{\gamma}_k \geq 1 - (1 + \mathrm{tr}(\Sigma\Phi_k))^{-\frac{1}{2}}.$$

Hence we obtain the relaxed constraint

$$\mathrm{tr}(\Sigma\Phi_k) \leq \left(\frac{1}{1-r_0}\right)^2 - 1. \tag{43}$$

Then we have a relaxed SDP optimization problem. The lower bound of optimal solution t to Problem 3 can be found by solving the following SDP problem.

Problem 4:

$$\min_{t,\{\Phi_k\}}\ t \tag{44}$$

$$\text{s.t. } \begin{bmatrix} \sum_{k=1}^{N} H_k\Sigma^{-1}H_k^\top - M_k & \mathrm{I}_n \\ \mathrm{I}_n & \frac{t}{N}\mathrm{I}_n \end{bmatrix} \geq 0, \tag{45}$$

$$\begin{bmatrix} \frac{1}{1-r_0}\left(\Phi_k+\Sigma^{-1}\right) & \Sigma^{-1}H_k^\top \\ H_k\Sigma^{-1} & M_k \end{bmatrix} \geq 0, \tag{46}$$

$$\mathrm{tr}(\Sigma\Phi_k) \leq \left(\frac{1}{1-r_0}\right)^2 - 1, \tag{47}$$

$$\Phi_k \geq 0, H_k\Sigma^{-1}H_k^\top \geq M_k \geq 0, \forall k. \tag{48}$$

The optimal solution to Problem 4 is the lower bound of the optimal solution to Problem 1. In particular, if the optimal $\Delta^\dagger$ to the Problem 4 satisfies (34), then $\Delta^* = \Delta^\dagger$.

Similarly we can find the upper bound of the optimal solution. By using [27, Lemma 2], we have the upper bound of $\bar{\gamma}_k$,

$$\bar{\gamma}_k \leq 1 - \exp\left(-\frac{1}{2}\mathrm{tr}(\Pi\Phi_k)\right).$$

The upper bound of optimal solution t to Problem 3 can be found by solving the following SDP problem.

Problem 5:

$$\begin{aligned} &\min_{t,\{\Phi_k\}}\ t \\ &\text{s.t. } \begin{bmatrix} \sum_{k=1}^{N} H_k\Sigma^{-1}H_k^\top - M_k & \mathrm{I}_n \\ \mathrm{I}_n & \frac{t}{N}\mathrm{I}_n \end{bmatrix} \geq 0, \\ &\quad \begin{bmatrix} \frac{1}{1-r_0}\left(\Phi_k+\Sigma^{-1}\right) & \Sigma^{-1}H_k^\top \\ H_k\Sigma^{-1} & M_k \end{bmatrix} \geq 0, \forall k, \\ &\quad \mathrm{tr}(\Sigma\Phi_k) \leq -2\ln(1-r_0), \forall k, \\ &\quad \Phi_k \geq 0, M_k \geq 0, \forall k. \end{aligned}$$

The tightness of the upper and lower bounds will be shown in the numerical examples later.

VI. Parameter Estimation Under Constrained Transmission Rate With Priori Knowledge

In this section we study the maximum *a posteriori* (MAP) estimation and the Bayes estimation problems when the priori information of θ is given. It is well known that if we have the priori statistical knowledge of θ, the ML estimation problem becomes the maximum *a posteriori* (MAP) estimation problem. More generally, to find the *a posteriori* distribution of the parameter, the ML estimation problem can be reformulated into the Bayes estimation problem. The simplicity of the MLE in (16) facilitate the two problems compared with the MLE in [29].

A. MAP Estimation

Assume we have the *a priori* distribution function $f_\theta(x)$ of θ, from Bayes' theorem,

$$f_\theta(x|\mathcal{Z}_N) = \frac{f_{\mathcal{Z}_N}(z_1,\ldots,z_N|\theta)f_\theta(x)}{f_{\mathcal{Z}_N}(z_1,\ldots,z_N)}. \tag{49}$$

Since $f_{Z_N}(z_1,\ldots,z_N)$ is independent of θ, the MAP estimator is thus given by

$$\begin{aligned} l_N^{\mathrm{MAP}}(\theta) &:= l_N(\theta) + \log\left(f_\theta(x)\right), \\ \hat{\theta}_N^{\mathrm{MAP}} &= \arg\max_\theta l_N^{\mathrm{MAP}}(\mathcal{Z}_N;\theta), \end{aligned}$$

where $l_N(\theta)$ is given in (13). Note that if $f_\theta(x)$ is log-concave then $l_N^{\mathrm{MAP}}(\theta)$ is concave due to the concavity of $l_N(\theta)$. In that case the global maximizer is easy to obtain via analytical or numerical methods.

Example 5 (Gaussian Case): Let $f_\theta(x) := \mathcal{N}(\tilde{\mu},\tilde{\Sigma})$ and set the gradient of $l_N^{\mathrm{MAP}}(\theta)$ to be zero, i.e.,

$$0 = \nabla\left(l_N^{\mathrm{MAP}}(\theta)\right) := \sum_{k=1}^{N} g_k(\theta) + \tilde{\Sigma}^{-1}(\tilde{\mu}-\theta)$$

Then we have

$$\hat{\theta}_N^{\mathrm{MAP}} = \left[\tilde{\Sigma}^{-1} + \sum_{k=1}^{N} \gamma_k H_k \Sigma^{-1} H_k^{\top} + (1-\gamma_k) H_k (\Sigma + \Delta_k)^{-1} H_k^{\top}\right]^{\dagger} \times \left[\tilde{\Sigma}^{-1}\tilde{\mu} + \sum_{k=1}^{N} \gamma_k H_k \Sigma^{-1} x_k + (1-\gamma_k) H_k (\Sigma + \Delta_k)^{-1} \tau_k\right]. \tag{50}$$

Note that if we have the non-informative *a priori* distribution for instance $\tilde{\Sigma} \to \infty$, then $\hat{\theta}_N^{\mathrm{MAP}} \to \hat{\theta}_N$. Particularly, when $N \to \infty$, the MAP estimator asymptotically converges to the MLE, i.e., $\hat{\theta}_N^{\mathrm{MAP}} \to \hat{\theta}_N$.

B. Bayes Estimation

The MAP estimation gives us the value where the *a posteriori* distribution function attains its maximum value. To obtain the *a posteriori* distribution function $f_\theta(x|\mathcal{Z}_N)$ in (49), we resort to solving the Bayes estimation problem. Usually, the most challenging part is to compute $f_{\mathcal{Z}_N}(z_1, \ldots, z_N) = \int f_{\mathcal{Z}_N}(z_1, \ldots, z_N|\theta = x) f_\theta(x) \mathrm{d}x$ where the integration may involve intractable computation. It is known that for a given likelihood function like $f_{\mathcal{Z}_N}(z_1, \ldots, z_N|\theta)$, a conjugate prior $f_\theta(x)$ is the *a priori* distribution which renders the *a posteriori* distribution $f_{\mathcal{Z}_N}(z_1, \ldots, z_N)$ own the same algebraic form as the *a priori* does. In practice one reasonably assumes a Gaussian *a priori* distribution function for θ since it is the conjugate prior for a Gaussian likelihood function.

One advantage of our scheduling policy is the computational simplicity of the resulting MLE. This is inherent in the derivation of the Bayes estimation. We can easily obtain the *a posteriori* distribution when the conjugate prior is used. If $f_\theta(x) := \mathcal{N}(\tilde{\mu}, \tilde{\Sigma})$, then we have

$$\begin{aligned} f_\theta(x|\mathcal{Z}_N) &= \frac{\prod_{k=1}^{N} f_{x_k}(x)^{\gamma_k} \Pr\{\gamma_k = 0\}^{1-\gamma_k} f_\theta(x)}{\int f_{\mathcal{Z}_N}(z_1, \ldots, z_N|\theta = x) f_\theta(x) \mathrm{d}x} \\ &= \mathcal{N}\left(\hat{\theta}_N^{\mathrm{MAP}}, P_N\right), \end{aligned} \tag{51}$$

where $\hat{\theta}_N^{\mathrm{MAP}}$ is given in (50) and P_N in (25). For a recursive MMSE estimator, we recommend [27] to the readers.

Remark 5: The Gaussian conjugate prior is assumed due to the Gaussian noise assumption. This is commonly adopted for simplicity despite any scheduling scheme is used. For other types of the distribution of noise, one may need to adjust the form of the policy in (3) to simplify the integration in (51).

VII. Numerical Simulations

In this section we present some numerical examples to show the main results and illustrate the performance of the proposed scheduling policy.

A. Some Properties of the MLE

We consider the following linear system model (1) with the parameter $\theta = [2\ 1]^{\top}$. 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

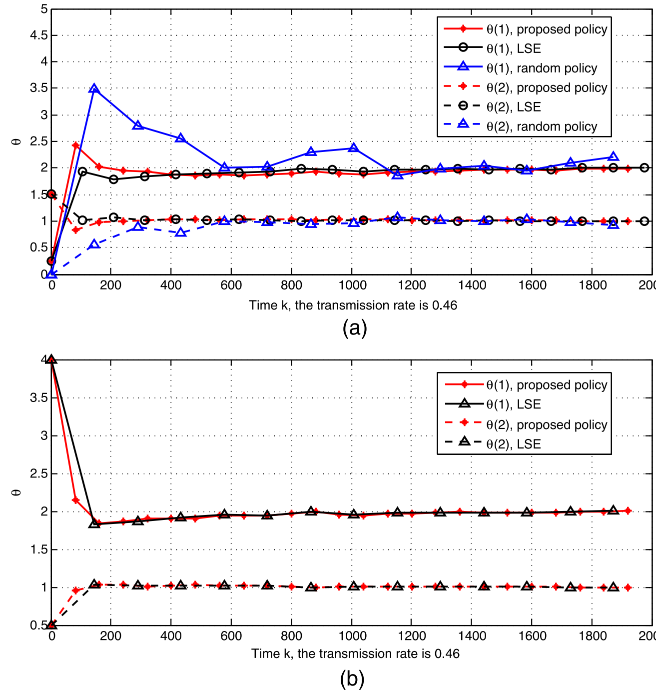


Fig. 2. (a) Comparison among the proposed MLE, the MLE with random drop policy and the LSE with full measurements. (b) MAP estimation under the proposed policy or with full measurements. The *a priori* distribution of θ is $\mathcal{N}([4, 0.5]^{\top}, \mathrm{I}_2)$.

the MLE estimator in (16) with $\tau_k = 2$ and $\Delta_k = 2, \forall k$. We choose the transmission rate constraint r_0 to be 0.46. As a reference we also plot the estimate of an LSE with full measurements and the estimate of an MLE with random transmission access at the same rate of 0.46. The MLE with random transmission 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. We can see that the convergence rate of the proposed MLE is much faster than that of the MLE with random transmission. Moreover, all the estimators are consistent due to the law of large number. In Fig. 2(b) we illustrate the MAP estimation. Assume the *a priori* distribution is $\mathcal{N}([4, 0.5]^{\top}, \mathrm{I}_2)$. It can be seen that the estimate is closer to the *a priori* mean and then converges to the true value after a long time. To see what the scheduling pattern is like, we zoom in on the time horizon in Fig. 2(a). The realization of γ_k within [450,500] is plotted in Fig. 3(b). It can be seen that the estimates of $\theta(2)$ given by the proposed MLE and LSE are quite close in Fig. 3(a) but the transmission rate has been reduced to 0.46. In Fig. 4 we plot the expected ML estimate versus time with 5000 simulation runs. The biasedness can be seen clearly when $k \leq 1000$. But the asymptotic unbiasedness is noticed when k becomes larger. In Fig. 5, the tradeoff between the transmission rate and the mean squared error (MSE) is shown. The trace of the MSE in dB is defined as $\mathrm{MSE[dB]} = 10\log_{10}(\mathrm{tr}(\mathbb{E}[\sqrt{N}(\hat{\theta}_N - \theta)\sqrt{N}(\hat{\theta}_N - \theta)^{\top}]))$. The time horizon is 500 and $\tau_k = 2, \forall k$. The different rates are chosen by adjusting $\Delta_k \in [0.01, 50], \forall k$.

B. Comparison With the MLE Under Deterministic Scheduler

In this subsection we aim to compare the proposed MLE with the MLE under the deterministic scheduler in [29]. In Fig. 6, we

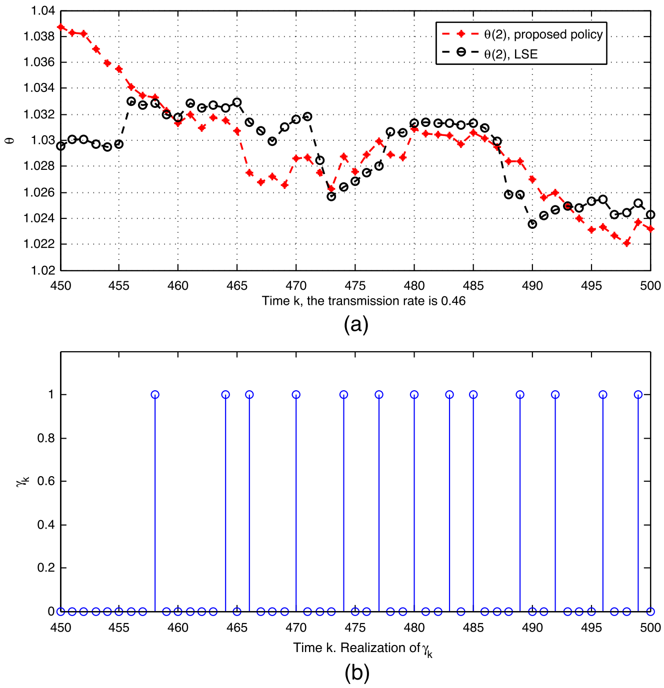


Fig. 3. (a) Zoom in on the horizon of [450,500]. Estimate of $\theta(2)$ given by the proposed MLE and LSE. (b) Realization of γ_k.

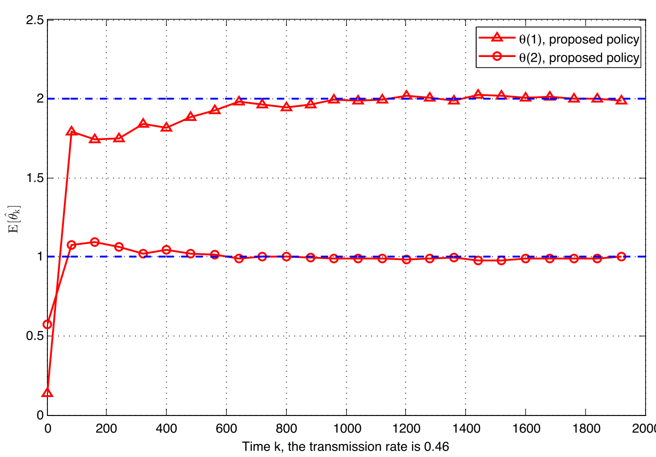


Fig. 4. Biasedness and asymptotic unbiasedness. The proposed MLE is biased at the beginning. The number of Monte Carlo simulation runs is 5000.

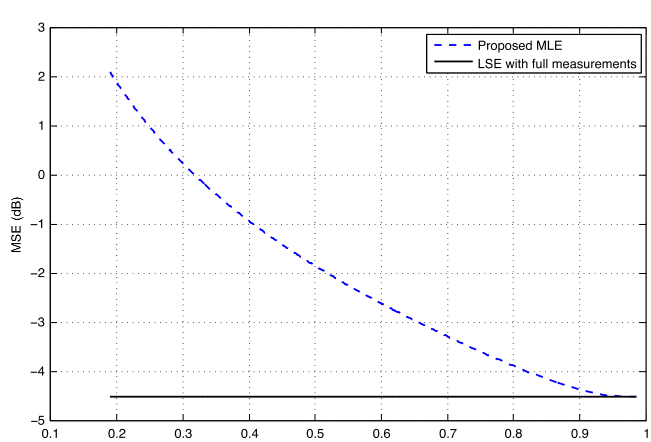


Fig. 5. Tradeoff between the transmission rate γ and the MSE in dB. The time horizon is 500.

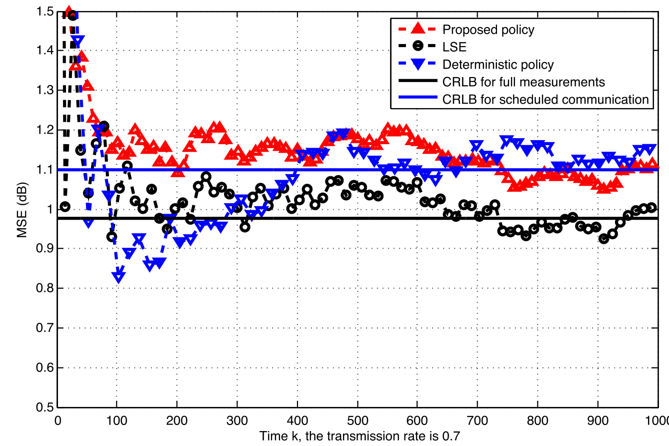


Fig. 6. Trace of the stationary error covariance for different estimators.

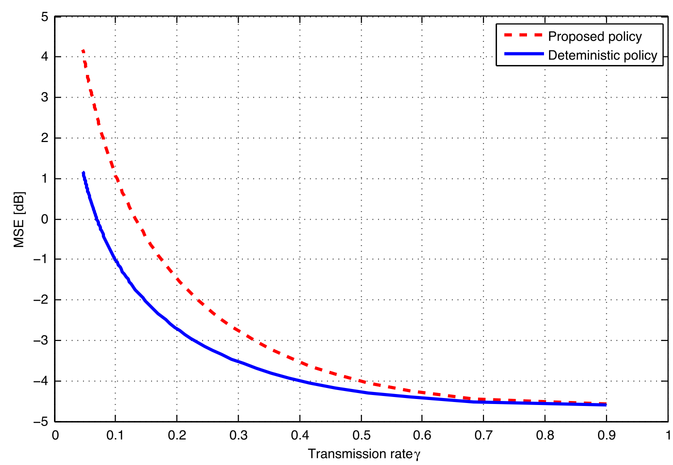


Fig. 7. Comparison with the deterministic scheduler.

compare the performance of the proposed MLE and the MLE with the deterministic policy in [26] under the same transmission rate 0.7. The thresholds in both policies are chosen to be $H_k^\top \theta$ for simplicity. The empirical asymptotic MSE under the proposed policy matches the theoretical MSE. In Fig. 7, we compare the performance of both policies under different transmission rates. To experimentally obtain an accurate transmission rate, we choose the time horizon to be 10000. It can be shown that the MSE under the proposed policy is larger since we add some random noise when the sensor is making a transmission decision. This reflects a tradeoff between the computation complexity and estimation performance. When the rate is larger, i.e. $\gamma > 0.5$, it is more likely to use the proposed policy since the performance gap is very small.

C. Parameter Design

In Section V, we relax Problem 1 into Problem 4 and 5 to find an upper bound and lower bound of the optimal objective function. If we denote $\rho(W_*^{-1}), \rho(W_-^{-1}), \rho(W_+^{-1})$ as the optimal objective functions by solving Problem 1, Problem 4 and Problem 5, respectively, it can be easily seen that

$$\rho\left(W_*^{-1}\right) - \rho\left(W_-^{-1}\right) \leq \rho\left(W_+^{-1}\right) - \rho\left(W_-^{-1}\right).$$

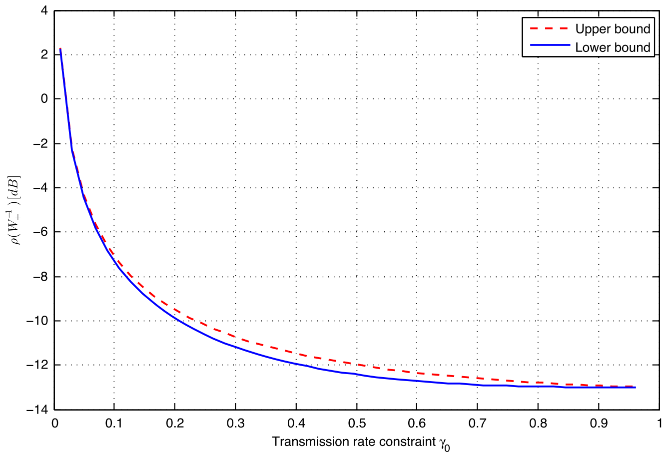


Fig. 8. Tightness of the upper and lower bounds of the optimal objection function.

In practice we can solve Problem 5 to obtain a suboptimal schedule satisfying the transmission rate constraint in (34). Thus the gap, $\rho(W_+^{-1}) - \rho(W_-^{-1})$, can quantify the performance difference between the optimal schedule and the suboptimal one. The tighter the gap is, the more closer $\rho(W_*^{-1})$ and $\rho(W_-^{-1})$ are to each other.

To illustrate the tightness of the upper and lower bounds, we choose the system parameters in (1) as follows,

$$H_k = \begin{bmatrix} 10 & 0.5 \\ 0 & 20 \end{bmatrix}, \forall k, \Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}.$$

Then we solve the two SDP problems in Problem 4 and 5 for each r_0 by using the `cvx` toolbox. Fig. 8 shows that the upper and lower bounds are tight for different transmission rate constraint, which means that the suboptimal schedule by solving Problem 5 can replace the optimal one very well.

VIII. Concluding Remarks

We have investigated the optimal parameter estimation under the transmission rate constraint in this work. We designed 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 obtained a closed-form data-driven MLE with a subset of measurements. We explicitly gave the CRLB of any unbiased estimator with the incomplete measurements. The asymptotic results show the consistency and asymptotic normality of the proposed MLE. The stationary error covariance of the MLE reaches the CRLB asymptotically. We also formulated 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. Due to the simplicity of the MLE, we also easily incorporated the prior knowledge of the parameter into the estimator.

One direction of the future work is to combine the temporal and spatial communication scheduling among a network of sensors. Our sensor scheduling policy is a temporal type which cannot avoid the packet collision and interference with other sensors. Taking the nodes in neighborhood into account while transmitting the important data is an interesting open problem.

Appendix

Proof of Lemma 3: Taking expectation over (14), we have

$$\begin{aligned}
&\mathbb{E}\left[g_k(\theta)\right] \\
&= \int\limits_{\{\gamma_k=1\}} H_k\Sigma^{-1}\left(x_k - H_k^\top\theta\right) f(x_k)\mathrm{d}x_k \\
&\quad + \int\limits_{\{\gamma_k=0\}} H_k(\Sigma+\Delta_k)^{-1}\left(\tau_k - H_k^\top\theta\right) f(x_k)\mathrm{d}x_k \\
&= \int\limits_{\mathbb{R}^m} H_k\Sigma^{-1}\left(H_k^\top\theta - x_k\right) f(x_k)\mathrm{d}x_k \\
&\quad + \int\limits_{\mathbb{R}^m} \left(H_k(\Sigma+\Delta_k)^{-1}\left(H_k^\top\theta - \tau_k\right)\right. \\
&\qquad \left. - H_k\Sigma^{-1}\left(H_k^\top\theta - x_k\right)\right) \\
&\qquad \times f(x_k)\exp\left(-\frac{1}{2}\|x_k-\tau_k\|^2_{\Delta_k^{-1}}\right)\mathrm{d}x_k \\
&= 0.
\end{aligned} \tag{52}$$

The first integral is 0 due to $\mathbb{E}[x_k] = H_k^\top\theta$. The second integral is also 0 from Lemma 1. ■

Proof of Proposition 1: To obtain the CRLB we need to first compute the Fisher information matrix $\mathfrak{I}_N := \mathbb{E}[\nabla(l_N(\theta))^\top\nabla(l_N(\theta))]$ where $\nabla(l_N(\theta))$ is given in (17). We have

$$\begin{aligned}
&\mathbb{E}\left[\nabla\left(l_N(\theta)\right)\nabla\left(l_N(\theta)\right)^\top\right] \\
&= \mathbb{E}\left[\left(\sum_{k=1}^N g_k(\theta)\right)^\top\left(\sum_{k=1}^N g_k(\theta)\right)\right] \\
&= \mathbb{E}\left[\sum_{k=1}^N g_k(\theta)g_k(\theta)^\top\right] + \mathbb{E}\left[\sum_{k\neq j}^N g_k(\theta)g_j(\theta)^\top\right] \\
&= \mathbb{E}\left[\sum_{k=1}^N g_k(\theta)g_k(\theta)^\top\right] + \sum_{k\neq j}^N \mathbb{E}\left[g_k(\theta)\right]\mathbb{E}\left[g_j(\theta)^\top\right] \qquad (53) \\
&= \sum_{k=1}^N \mathbb{E}\left[g_k(\theta)g_k(\theta)^\top\right]. \qquad (54)
\end{aligned}$$

The latter term in (53) results from the fact that $\{g_k(\theta)\}$ is an independent random process. Then from (3), (14) and (54), we have

$$\begin{aligned}
&\mathbb{E}\left[\nabla\left(l_N(\theta)\right)\nabla\left(l_N(\theta)\right)^\top\right] \\
&= \sum_{k=1}^N H_k\Sigma^{-1}\mathbb{E}_{\{\gamma_k=1\}} \\
&\quad \times \left[\left(x_k - H_k^\top\theta\right)\left(x_k - H_k^\top\theta\right)^\top\right]\left(H_k\Sigma^{-1}\right)^\top \\
&\quad + H_k(\Sigma+\Delta_k)^{-1}\mathbb{E}_{\{\gamma_k=0\}}\left[\left(\tau_k - H_k^\top\theta\right)\left(\tau_k - H_k^\top\theta\right)^\top\right] \\
&\quad \times \left(H_k(\Sigma+\Delta_k)^{-1}\right)^\top.
\end{aligned} \tag{55}$$

From (6) and Lemma 1, the first expectation term in (55) is given by

$$\begin{aligned}
&\mathbb{E}_{\{\gamma_k=1\}}\left[\left(x_k - H_k^\top\theta\right)\left(x_k - H_k^\top\theta\right)^\top\right] \\
&= \int_{\mathbb{R}^m} \left(x_k - H_k^\top\theta\right)\left(x_k - H_k^\top\theta\right)^\top f(x_k)\mathrm{d}x_k \\
&\quad - \int_{\mathbb{R}^m} \left(x_k - H_k^\top\theta\right)\left(x_k - H_k^\top\theta\right)^\top f(x_k) \\
&\quad \times \exp\left(-\frac{1}{2}\|x_k - \tau_k\|^2_{\Delta_k^{-1}}\right)\mathrm{d}x_k \\
&= \Sigma - \alpha_k\left[\Omega_k + \left(\omega_k - H_k^\top\theta\right)\left(\omega_k - H_k^\top\theta\right)^\top\right]. \quad (56)
\end{aligned}$$

The second expectation term in (55) is given by

$$\begin{aligned}
&\mathbb{E}_{\{\gamma_k=0\}}\left[\left(\tau_k - H_k^\top\theta\right)\left(\tau_k - H_k^\top\theta\right)^\top\right] \\
&\qquad = \alpha_k\left(\tau_k - H_k^\top\theta\right)\left(\tau_k - H_k^\top\theta\right)^\top. \quad (57)
\end{aligned}$$

From (55), (56), (57) and Lemma 2, we have

$$\begin{aligned}
&\mathbb{E}\left[\nabla\left(l_N(\theta)\right)\nabla\left(l_N(\theta)\right)^\top\right] \\
&\qquad = \sum_{k=1}^N H_k\Sigma^{-1}H_k^\top - \alpha_k H_k\Sigma^{-1}\Omega_k\Sigma^{-1}H_k^\top. \quad (58)
\end{aligned}$$

Therefore, we have the CRLB for any unbiased estimator

$$\mathrm{Var}\left(\theta_N^u\right) \geq \mathbb{E}\left[\nabla\left(l_N(\theta)\right)\nabla\left(l_N(\theta)\right)^\top\right]^{-1} = \mathfrak{I}_N^{-1}.$$

■

Proof of Theorem 3:

1) Notice that $\hat{\theta}_N - \theta = S_N^{-1}T_N$, where

$$S_N := \frac{1}{N}\sum_{k=1}^N h_k(\theta), T_N := \frac{1}{N}\sum_{k=1}^N g_k(\theta).$$

The functions $g_k(\theta)$ and $h_k(\theta)$ are defined in (14) and (15). We now show that $S_N \xrightarrow{a.s.} \lim_{N\longrightarrow+\infty}\mathbb{E}[S_N]$ and $T_N \xrightarrow{a.s.} \lim_{N\rightarrow+\infty}\mathbb{E}[T_N]$. From Lemma 2, we have

$$\begin{aligned}
\mathbb{E}\left[h_k(\theta)\right] &= (1-\alpha_k)H_k\Sigma^{-1}H_k^\top + \alpha_k H_k(\Sigma+\Delta_k)^{-1}H_k^\top, \\
\mathbb{E}\left[h_k(\theta)h_k(\theta)^\top\right] &= \alpha_k(1-\alpha_k)\left(H_k\Sigma^{-1}H_k^\top H_k\Sigma^{-1}H_k^\top\right. \\
&\quad \left.+ H_k(\Sigma+\Delta_k)^{-1}H_k^\top H_k(\Sigma+\Delta_k)^{-1}H_k^\top\right).
\end{aligned}$$

Since γ_k is independently distributed and $\sup_k\|\mathbb{E}[h_k(\theta)h_k(\theta)^\top]\| \leq \infty$, from the Rajchman's strong law of large numbers (SLLN) [32, Theorem 5.1.2] we conclude that

$$S_N \xrightarrow{a.s.} \lim_{N\rightarrow+\infty}\frac{1}{N}\sum_{k=1}^N \mathbb{E}\left[h_k(\theta)\right] = W^{-1}. \qquad (59)$$

From (52) we have $\mathbb{E}[g_k(\theta)] = 0$. From (56) and (57), we can see that $\sup_k\|\mathbb{E}[g_k(\theta)g_k(\theta)^\top]\| \leq \infty$. Now we have

$$T_N \xrightarrow{a.s.} 0$$

Since $\lim_{N\rightarrow+\infty}\frac{1}{N}\sum_{k=1}^N \mathbb{E}[\|h_k(\theta)\|] < \infty$, from the continuous mapping theorem [33, Theorem 2.3], we have that

$$\hat{\theta}_N - \theta = S_N^{-1}T_N \xrightarrow{a.s.} 0,$$

which proves the consistency of $\hat{\theta}_N$.

2) By the dominated convergence theorem [34], it is straightforward to obtain the result.

3) Pick an arbitrary $y \in \mathbb{R}^n$ and denote $\tilde{g}_k(\theta) = y^\top g_k(\theta)$. We shall show the Lyapunov condition [34] holds and prove the asymptotic normality of T_N. For $\varepsilon > 0$, there exists an upper bound $M_1 > 0$ such that

$$\begin{aligned}
U_1 &= \frac{1}{N}\sum_{k=1}^N \mathbb{E}\left[\|\tilde{g}_k(\theta)\|^{2+\varepsilon}\right] = \frac{1}{N}\sum_{k=1}^N \mathbb{E}\left[\|y^\top g_k(\theta)\|^{2+\varepsilon}\right] \\
&\leq \frac{2^{1+\varepsilon}}{N}\sum_{k=1}^N \mathbb{E}\left[\|\gamma_k y^\top H_k\Sigma^{-1}\nu_k\|^{2+\varepsilon}\right] \\
&\quad + \frac{2^{1+\varepsilon}}{N}\sum_{k=1}^N \mathbb{E}\left[\|(1-\gamma_k)y^\top H_k(\Sigma+\Delta_k)^{-1}\right. \\
&\qquad \left.\left(\tau_k - H_k^\top\theta\right)\|^{2+\varepsilon}\right] \\
&\leq \frac{2^{1+\varepsilon}}{N}\sum_{k=1}^N \mathbb{E}\left[\|y^\top H_k\Sigma^{-1}\nu_k\|^{2+\varepsilon}\right] \\
&\quad + \frac{2^{1+\varepsilon}}{N}\sum_{k=1}^N \mathbb{E}\left[\|y^\top H_k(\Sigma+\Delta_k)^{-1}(\tau_k - H_k^\top\theta)\|^{2+\varepsilon}\right] \\
&\leq M_1 \leq \infty.
\end{aligned}$$

The first inequality comes from the c_r inequality [35, Chapter 6, Theorem 4]. The boundness is due to $\sup_k\|H_k\| \leq \infty$ and $\sup_k\|\tau_k\| \leq \infty$. On the other hand, there exists an upper bound M_2 such that

$$U_2 = \frac{1}{N}\sum_{k=1}^N \mathbb{E}\left[\|\tilde{g}_k(\theta)\|^2\right] = \frac{1}{N}y^\top\mathfrak{I}_N y \leq M_2 \leq \infty.$$

Then we have

$$\frac{U_1}{N^{\frac{\varepsilon}{2}}U_2^{1+\frac{\varepsilon}{2}}} \rightarrow 0$$

which shows the Lyapunov condition is satisfied. Since Assumption 1 ensures that W exists, applying the Lindeberg-Feller central limit theorem [34, Theorem 7.3.1] and from (58) we have

$$\sqrt{N}T_N \xrightarrow{d} \mathcal{N}(0, W^{-1}).$$

Note that

$$\sqrt{N}(\hat{\theta}_N - \theta) = S_N^{-1}(\sqrt{N}T_N).$$

From (59) and the Slutsky's theorem [35, Chapter 5, Theorem 7], we have

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \mathcal{N}(0, W^{-1}).$$

4) From (11) and the SLLN, we immediately have (30) which completes the proof. ■

Proof of Theorem 4: The consistency and asymptotic normality are not affected by the value of τ and their validity follows from Theorem 3. The stationary covariance W_+^{-1} is different and needs to be revisited. From (27), we can rewrite the new Fisher information matrix in case that $\tau_k = H_k^\top \theta_{k-1}$,

$$\mathfrak{I}_N^* := \sum_{k=1}^{N} H_k \Sigma^{-1} H_k^\top - \alpha_k^* H_k \Sigma^{-1} \Omega_k \Sigma^{-1} H_k^\top, \quad (60)$$

$$\alpha_k^* = \det\left(\mathbf{I}_n + \Sigma \Delta_k^{-1}\right)^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2}\left\|H_k^\top \theta - H_k \hat{\theta}_{k-1}\right\|^2_{\Delta_k + \Sigma^{-1}}\right). \quad (61)$$

Since $\hat{\theta}_N^* \xrightarrow{a.s.} \theta$, we have $\tau_k \xrightarrow{a.s.} H_k^\top \theta$. Two immediate results are $\alpha_k^* \xrightarrow{a.s.} \det(\mathbf{I}_n + \Sigma\Delta_k^{-1})^{-\frac{1}{2}}$ and $\frac{1}{N}\mathfrak{I}_N^* \xrightarrow{a.s.} W_+$ according to the continuous mapping theorem. Together with (30) we have (32). Due to the asymptotic efficiency of the MLE, we conclude the stationary covariance is W_+^{-1}. ∎

Proof of Lemma 4: To facilitate the proof, we introduce some definitions and notations. Denote the operator $L(\Delta) := W^{-1}(\Delta)$ where W^{-1} is given in (29). The increasing operator monotonicity is asserted if and only if $L(\Delta_1) > L(\Delta_2)$ for any $\Delta_1 > \Delta_2$, where $\Delta_i := \{\Delta_k^i\}, i = 1, 2, k = 1, \ldots, N$. We mean the sequence order $\Delta_1 > \Delta_2$ by $\Delta_k^1 - \Delta_k^2 > 0$, for any k.

It is straightforward to show that $L(\Delta)$ is operator monotonically increasing. Now we shall prove the lemma by contradiction. If we have the optimal solution (τ^*, Δ^*) to the Problem 1 where $\exists k, \tau_k^* \neq H_k^\top \theta$, there exists another solution (τ', Δ') where $\Delta_k' < \Delta_k^*$ and $\tau_k' = H_k^\top \theta$ for the indices $k \in \{k | \tau_k^* \neq H_k^\top \theta\}$ and $\Delta_k' = \Delta_k^*$ for the indices $k \in \{k | \tau_k^* = H_k^\top \theta\}$ such that

$$\det\left(\mathbf{I}_n + \Sigma(\Delta_k^*)^{-1}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left\|H_k^\top \theta - \tau_k^*\right\|^2_{\Delta_k^* + \Sigma^{-1}}\right)$$
$$= \det\left(\mathbf{I}_n + \Sigma\left(\Delta_k'\right)^{-1}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left\|H_k^\top \theta - H_k^\top \theta\right\|^2_{\Delta_k' + \Sigma^{-1}}\right)$$

due to the continuity of $\det(\cdot)$. Thus both solutions lead to the same $\bar{\gamma}_N$. Due to the operator monotonicity of $L(\Delta)$, we have $L(\Delta_k') < L(\Delta_k^*)$, thus $\rho(L(\Delta_k')) < \rho(L(\Delta_k^*))$ which contradicts the optimality assumption of (τ^*, Δ^*). This completes the proof. ∎

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[2] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[3] J. Li and G. AlRegib, "Rate-constrained distributed estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1634–1643, May 2007.

[4] Z. Duan, V. P. Jilkov, and X. R. Li, "State estimation with quantized measurements: Approximate MMSE approach," in *Proc. Int. Conf. Inf. Fusion*, 2008, pp. 1–6.

[5] M. Fu and C. E. De Souza, "State estimation using quantized measurements," presented at the IFAC, 2008.

[6] B. I. Godoy, G. C. Goodwin, J. C. Agüero, D. Marelli, and T. Wigren, "On identification of FIR systems having quantized output data," *Automatica*, vol. 47, no. 9, pp. 1905–1915, 2011.

[7] D. Marelli, K. You, and M. Fu, "Identification of ARMA models using intermittent and quantized output observations," *Automatica*, vol. 49, no. 2, pp. 360–369, 2013.

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

[9] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proc. IEEE*, vol. 95, no. 1, pp. 163–187, Jan. 2007.

[10] M. Huang and S. Dey, "Stability of kalman filtering with Markovian packet losses," *Automatica*, vol. 43, no. 4, pp. 598–607, 2007.

[11] Y. Mo and B. Sinopoli, "Kalman filtering with intermittent observations: Tail distribution and critical value," *IEEE Trans. Autom. Control*, vol. 57, no. 3, pp. 677–689, Mar. 2012.

[12] T. Sui, K. You, M. Fu, and D. Marelli, "Stability of MMSE state estimators over lossy networks using linear coding," *Automatica*, vol. 51, pp. 167–174, 2015.

[13] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proc. IEEE*, vol. 95, no. 1, p. 138, Jan. 2007.

[14] L. Y. Wang, J.-F. Zhang, and G. G. Yin, "System identification using binary sensors," *IEEE Trans. Autom. Control*, vol. 48, no. 11, pp. 1892–1907, Nov. 2003.

[15] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, Jul. 2006.

[16] K. You, "Recursive algorithms for parameter estimation with adaptive quantizer," *Automatica*, vol. 52, pp. 192–201, 2015.

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

[18] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises," *J. Wireless Commun. Netw.*, vol. 2005, no. 4, pp. 473–482, 2005.

[19] J.-Y. Wu, Q.-Z. Huang, and T.-S. Lee, "Minimal energy decentralized estimation via exploiting the statistical knowledge of sensor noise variance," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2171–2176, May 2008.

[20] J. Li and G. AlRegib, "Distributed estimation in energy-constrained wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3746–3758, Oct. 2009.

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

[22] D. Han, P. Cheng, J. Chen, and L. Shi, "An online sensor power schedule for remote state estimation with communication energy constraint," *IEEE Trans. Autom. Control*, vol. 59, no. 7, pp. 1942–1947, Jul. 2014.

[23] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[24] L. Shi, K. H. Johansson, and L. Qiu, "Time and event-based sensor scheduling for networks with limited communication resources," in *Proc. World Congr. IFAC*, 2011, vol. 18, no. 1, pp. 13 263–13 268.

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

[26] K. You and L. Xie, "Kalman filtering with scheduled measurements," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1520–1530, Mar. 2013.

[27] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Trans. Autom. Control*, 2015, to be published.

[28] D. Shi, T. Chen, and L. Shi, "Event-triggered maximum likelihood state estimation," *Automatica*, vol. 50, no. 1, pp. 247–254, 2014.

[29] K. You, L. Xie, and S. Song, "Asymptotically optimal parameter estimation with scheduled measurements," *IEEE Trans. Signal Process.*, vol. 61, no. 14, pp. 3521–3531, Jul. 2013.

[30] L. Ljung, *Systems Identification-Theory for the User*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

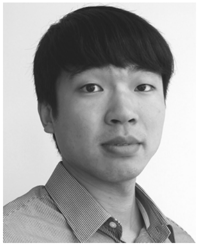
[31] M. Green and J. B. Moore, "Persistence of excitation in linear systems," in *Proc. Amer. Control Conf.*, 1985, pp. 412–417.

[32] K. L. Chung, *A Course in Probability Theory*. San Diego, CA, USA: Academic Press, 2001.

[33] A. W. Van der Vaart, *Asymptotic Statistics*. Cambridge, U.K.: Cambridge Univ. Press, 2000, vol. 3.

[34] R. B. Ash and C. Doleans-Dade, *Probability and Measure Theory*. San Diego, CA, USA: Academic Press, 2000.

[35] G. G. Roussas, *An Introduction to Measure-Theoretic Probability*. San Diego, CA, USA: Academic Press, 2004.



Duo Han was born in Heilongjiang, China, in 1990. He received his B.Eng. (Hons) degree from City University of Hong Kong, Hong Kong in 2011 and his Ph.D. degree in Electronic and Computer Engineering from the Hong Kong University of Science and Technology, Hong Kong in 2015. He was a visiting student in Control and Dynamical Systems at the California Institute of Technology, Pasadena, CA, USA in 2014. He is currently a research fellow at the School of Electrical and Electronic Engineering of Nanyang Technological University, Singapore. His reseach interests include cyber-physical systems, wireless sensor scheduling and event-based state estimation.

Keyou You was born in Jiangxi Province, China, in 1985. He received the B.S. degree in statistical science from Sun Yat-sen University, Guangzhou, China, in 2007 and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2012.

From June 2011 to June 2012, he was with the Sensor Network Laboratory at NTU as a Research Fellow. Since July 2012, he has been with the Department of Automation, Tsinghua University, China as an Assistant Professor. He held visiting positions at The Hong Kong University of Science and Technology, The University of Melbourne, Politecnico di Torino and etc. His current research interests include control and estimation of networked systems, distributed algorithms, and sensor network.

Dr. You received the Guan Zhaozhi best paper award at the 29th Chinese Control Conference in 2010, and a CSC-IBM China Faculty Award in 2014. He was selected to the national "1000-Youth Talent Program" of China in 2014.

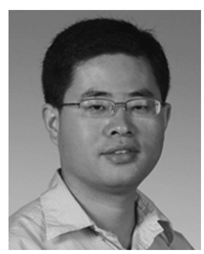
Lihua Xie (S'91–M'92–SM'97–F'07) received the B.E. and M.E. degrees in electrical engineering from Nanjing University of Science and Technology in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Newcastle, Australia, in 1992. Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a professor. He served as the Head of Division of Control and Instrumentation from July 2011 to June 2014. He held teaching appointments in the Department of Automatic Control, Nanjing University of Science and Technology from 1986 to 1989.

Dr. Xie's research interests include robust control and estimation, networked control systems, multi-agent networks, and unmanned systems. He has served as an Editor-in-Chief of Unmanned Systems, an editor of IET Book Series in Control and an Associate Editor of a number of journals including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, Automatica, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II.

Dr. Xie is a Fellow of IEEE and Fellow of IFAC.

Junfeng Wu received the B.Eng. degree from the Department of Automatic Control, Zhejiang University, Hangzhou, China, in 2009 and the Ph.D. degree in Electrical and Computer Engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2013. From September to December 2013, he was a research associate in Department of Electronic and Computer Engineering, the Hong Kong University of Science and Technology, Hong Kong. Dr. Wu is currently a postdoctoral researcher at ACCESS (Autonomic Complex Communication nEtworks, Signals and Systems) Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden. His research interests include networked control systems, state estimation, and wireless sensor networks, multi-agent systems.

Dr. Wu received the Guan Zhao-Zhi best paper award at the 34th Chinese Control Conference in 2015.



Ling Shi received the B.S. degree in electrical and electronic engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2002 and the Ph.D. degree in control and dynamical systems from California Institute of Technology, Pasadena, CA, USA, in 2008. He is currently an associate professor at the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. His research interests include networked control systems, wireless sensor networks, event-based state estimation and sensor scheduling, and smart energy systems. He has been serving as a subject editor for International Journal of Robust and Nonlinear Control from 2015. He also served as an associate editor for a special issue on Secure Control of Cyber Physical Systems in the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS IN 2015.