The Annals of Statistics
2012, Vol. 40, No. 4, 2239–2265
DOI: 10.1214/12-AOS1035

ASYMPTOTICALLY OPTIMAL PARAMETER ESTIMATION UNDER COMMUNICATION CONSTRAINTS

BY GEORGIOS FELLOURIS

University of Southern California

A parameter estimation problem is considered, in which dispersed sensors transmit to the statistician partial information regarding their observations. The sensors observe the paths of continuous semimartingales, whose drifts are linear with respect to a common parameter. A novel estimating scheme is suggested, according to which each sensor transmits only one-bit messages at stopping times of its local filtration. The proposed estimator is shown to be consistent and, for a large class of processes, asymptotically optimal, in the sense that its asymptotic distribution is the same as the exact distribution of the optimal estimator that has full access to the sensor observations. These properties are established under an asymptotically low rate of communication between the sensors and the statistician. Thus, despite being asymptotically efficient, the proposed estimator requires minimal transmission activity, which is a desirable property in many applications. Finally, the case of discrete sampling at the sensors is studied when their underlying processes are independent Brownian motions.

1. Introduction. Consider a number of dispersed sensors, each one of which observes the path of a real-valued stochastic process. The joint distribution of these processes is assumed to belong to some parametric family. The goal is to estimate the unknown parameter at a central location (*fusion center*) that receives information from all sensors.

When the sensors transmit their complete observations to the fusion center, we have a classical (*centralized*) parameter estimation problem. However, the fusion center often does not have full access to the sensor observations due to practical considerations, such as limited communication bandwidth. These communication constraints are present in applications such as mobile and wireless communication, data fusion, environmental monitoring and distributed surveillance, in which it is crucial to minimize the congestion in the network and the computational burden at the fusion center (see, e.g., Foresti et al. [6]).

Under this setup, which is often called *decentralized*, each sensor needs to transmit a small number of bits per communication to the fusion center and it is clear that the classical (centralized) statistical techniques are no longer applicable. As

Received February 2011; revised July 2012.

MSC2010 subject classifications. Primary 62L12, 62F30; secondary 62F12, 62M05, 62M09.

Key words and phrases. Asymptotic optimality, communication constraints, decentralized estimation, quantization, random sampling, sequential estimation, semimartingale.

a result, there has been a great interest in decentralized formulations of statistical problems (see, e.g., the review papers by Viswanathan and Varshney [24], Blum et al. [1], Han and Amari [10] and Veeravalli [23]).

Parameter estimation under a decentralized setup has been studied extensively using information-theoretic techniques. More specifically, it is often assumed that there are two correlated sensors, each of which observes a sequence of independent and identically distributed (i.i.d.), finite-valued random variables whose joint probability mass function is determined by the unknown parameter. The sensors are then required to transmit to the fusion center messages that belong to alphabets of smaller size than those of the original observations. The review paper by Han and Amari [9] describes in detail the main advances in this line of research. On the other hand, Luo [15] and Xiao and Luo [25] considered an arbitrary number of independent sensors that take i.i.d. observations with a common mean, which is the unknown parameter. Assuming that the parameter space and the support of the noise distribution are both compact intervals, they constructed decentralized estimating schemes that require the transmission of a small number of bits per communication.

In all the above papers, the sensors collect i.i.d. observations at a sequence of discrete times and transmit a small number of bits to the fusion center at every such sampling time. Moreover, even under an asymptotically large horizon of observations, the resulting estimators have larger mean square errors than the corresponding optimal centralized estimators, which have full access to the sensor observations.

In this paper the goal is to construct a decentralized estimating scheme that requires minimal communication activity from the sensors *and* achieves asymptotically the mean square error of the optimal centralized estimator, under a general statistical model for the sensor observations. In particular, we assume that the sensors observe the paths of continuous semimartingales whose drifts are linear with respect to the unknown parameter.

The centralized version of this problem is well understood. For Gaussian processes with independent increments, the fixed-horizon maximum likelihood estimator (MLE) was studied by Grenander [8] and Striebel [22]. Brown and Hewitt [2] proved that the MLE is consistent and asymptotically normal for stationary and ergodic time-homogeneous diffusions. Feigin [4] established the same properties for more general diffusions, assuming that the score process is a martingale. Liptser and Shiryaev ([13], pages 225–236) studied the MLE for a diffusion-type process and computed its bias and variance in the Ornstein–Uhlenbeck case. For a diffusion-type process with linear drift with respect to the unknown parameter, Liptser and Shiryaev [13], pages 244–248, and earlier Novikov [18], suggested a *sequential* version of the MLE and proved that it is unbiased and that it attains a prescribed accuracy. In the particular case of a square root diffusion, Brown and Hewitt [3] suggested an alternative sequential estimator with similar optimality properties. Melnikov and Novikov [17] and Galtchouk and Konev [7] studied least-squares sequential estimators that attain a prescribed accuracy in a multidimensional semimartingale regression model, generalizing in this way the results of Novikov [18]. We refer to Kutoyants [12] and Rao [19] for exhaustive references in the statistical inference of diffusion and diffusion-type processes.

Apart from the statistical model for the sensor observations, our work differs from previous approaches in some other important aspects as well. First of all, we do not assume that the frequency with which a sensor transmits its messages to the fusion center (*communication rate*) is the same as the frequency with which it collects its local observations (*sampling rate*). Instead, we assume that the sensors observe their underlying processes continuously, but communicate with the fusion center at discrete times. Therefore, in our context, the incurred loss of information is not only due to the quantization of sensor observations, but also due to the discrete transmission of messages to the fusion center in comparison to the continuous flow of information at the sensors.

Moreover, we do not require that the sensors communicate with the fusion center at deterministic and equidistant times. Instead, we allow each sensor to transmit its messages to the fusion center at random times that are triggered by its local observations. In particular, we propose a communication scheme according to which the sensors transmit only *one-bit* messages at first exit times of appropriate, locally-observed statistics (see Rabi et al. [20] and Fellouris and Moustakides [5] for similar communication schemes in different decentralized problems). Based on this communication scheme, we construct an estimator that is always consistent, even when the sensor processes are dependent.

However, the main result of this paper is that, in certain cases, the asymptotic distribution of the proposed estimator is the same as the exact distribution of the corresponding optimal centralized estimator. In particular, this holds when the sensor processes are arbitrary, orthogonal continuous semimartingales, as well as when they are correlated Gaussian processes with independent increments.

More importantly, these asymptotic properties are established as the horizon of observations goes to infinity and as the rate of communication between sensors and the fusion center goes to *zero*. Thus, although the proposed estimator is statistically efficient, it requires minimal communication activity from the sensors, which is a very desirable property in applications with severe communication constraints.

Finally, we consider in more detail the special case in which the sensors observe independent Brownian motions, since the tractability of this model allows us to obtain additional insight regarding the suggested estimating scheme. In this context, we also consider the case of discrete sampling, where the sensors do not observe their underlying processes continuously, but at a sequence of discrete times. It is shown that the proposed estimator remains consistent for any fixed sampling frequency, as long as the sensors have an asymptotically low rate of communication with the fusion center. However, asymptotic optimality does require a sufficiently high sampling rate, which we determine as a function of the communication rate and the observation horizon.

The rest of the paper is organized as follows: in Section 2 we formulate the problem under consideration. In Section 3 we specify the proposed estimating scheme and analyze its asymptotic properties. In Section 4 we focus on the special case that the sensors observe independent Brownian motions. We conclude in Section 5.

2. Problem formulation.

In what follows, we denote by i the generic sensor, where $i = 1, \dots, K$. We assume that sensor i observes the path of a continuous stochastic process $Y^i = \{Y^i_t\}_{t \geq 0}$ and is able to compute any statistic that is adapted to the filtration generated by Y^i.

In this section we specify the dynamics of $(Y^1, \dots, Y^K)$ under a family of probability measures $\{\mathsf{P}_\lambda, \lambda \in \mathbb{R}\}$, we review standard results regarding the centralized estimation of the unknown parameter λ and we define the notion of an (asymptotically optimal) decentralized estimator.

2.1. *Statistical model.* Let $(Y^1, \dots, Y^K)$ be the coordinate process on the canonical space of continuous functions $(\Omega, \mathcal{F})$, where $\Omega := \mathbb{C}[0, \infty)^K$ and $\mathcal{F} := \mathcal{B}(\Omega)$ is the associated Borel σ-algebra. We denote by $\{\mathcal{F}^i_t\}$ the right-continuous version of the natural filtration generated by Y^i and by $\{\mathcal{F}_t\}$ the corresponding global filtration

$$\mathcal{F}^i_t := \mathcal{C}^i_{t+}, \qquad \mathcal{C}^i_t := \sigma\big(Y^i_s; 0 \leq s \leq t\big), \tag{2.1}$$

$$\mathcal{F}_t := \mathcal{C}_{t+}, \qquad \mathcal{C}_t := \sigma\big(Y^i_s; 0 \leq s \leq t, 1 \leq i \leq K\big). \tag{2.2}$$

Let also P_0 be a probability measure on $(\Omega, \mathcal{F})$ so that

$$Y^i \in \mathcal{M}_0 \qquad \forall 1 \leq i \leq K,$$

where $\mathcal{M}_0$ is the class of continuous P_0-local martingales that start from 0.

For every $1 \leq i, j \leq K$, we denote by $\langle Y^i, Y^j \rangle$ the quadratic covariation of Y^i and Y^j and we assume that X^i is an $\{\mathcal{F}^i_t\}$-progressively measurable process so that

$$\mathsf{P}_0\left(\sum_{i=1}^K \int_0^t |X^i_s|^2 \, \mathrm{d}\langle Y^i, Y^i \rangle_s < \infty\right) = 1 \qquad \forall 0 \leq t < \infty. \tag{2.3}$$

Then, we can define the stochastic integral

$$B_t := \sum_{i=1}^K \int_0^t X^i_s \, \mathrm{d}Y^i_s, \qquad t \geq 0, \tag{2.4}$$

and we denote by A its quadratic variation, that is,

$$A_t := \langle B, B \rangle_t = \sum_{i=1}^K \sum_{j=1}^K \int_0^t X^i_s X^j_s \, \mathrm{d}\langle Y^i, Y^j \rangle_s, \qquad t \geq 0. \tag{2.5}$$

Moreover, we assume that the Novikov-type condition:

$$\mathsf{E}_0\big[e^{(\lambda^2/2)A_t}\big] < \infty \qquad \forall 0 \le t < \infty \tag{A1}$$

is satisfied for every $\lambda \neq 0$, which allows us to define for every $\lambda \neq 0$ the probability measure P_λ in the following way:

$$\frac{d\mathsf{P}_\lambda}{d\mathsf{P}_0}\bigg|_{\mathcal{F}_t} := e^{\lambda B_t - (\lambda^2/2)A_t} \qquad \forall 0 \le t < \infty. \tag{2.6}$$

Then, if we denote by $\mathcal{M}_\lambda$ the class of continuous P_λ-local martingales that start from 0, Girsanov's theorem (see [21], page 331) implies that

$$N^i := Y^i - \langle Y^i, \lambda B\rangle \in \mathcal{M}_\lambda \qquad \forall i = 1, \dots, K \tag{2.7}$$

and, consequently, $\langle N^i, N^j\rangle = \langle Y^i, Y^j\rangle$ for every $i \neq j$. Therefore, from (2.4) and (2.7) it follows that under P_λ

$$Y_t^i = \lambda \sum_{j=1}^{K} \int_0^t X_s^j \, d\langle Y^i, Y^j\rangle_s + N_t^i, \qquad t \ge 0, 1 \le i \le K. \tag{2.8}$$

2.2. *The parameter estimation problem.* The goal is to estimate the unknown parameter λ using the information that is being transmitted from the sensors to the fusion center. The flow of this information can be described by a sub-filtration of $\{\mathcal{F}_t\}$ and is determined by the *communication scheme* that is chosen by the statistician.

Let $\{\mathcal{G}_t\} \subset \{\mathcal{F}_t\}$ be the fusion center filtration. We will say that:

(a) $(\phi_t)_{t>0}$ is a *fixed-horizon*, $\{\mathcal{G}_t\}$-adapted estimator of λ, if ϕ_t is a $\mathcal{G}_t$-measurable statistic for every $t > 0$.

(b) $(T_\gamma, \phi_\gamma)_{\gamma>0}$ is a *sequential*, $\{\mathcal{G}_t\}$-adapted estimator of λ, if $(T_\gamma)_{\gamma>0}$ is an increasing family of $\{\mathcal{G}_t\}$-stopping times and ϕ_γ a $\mathcal{G}_{T_\gamma}$-measurable statistic for every $\gamma > 0$.

We will say that a $\{\mathcal{G}_t\}$-adapted estimator, either fixed-horizon or sequential, is *decentralized*, when the fusion center filtration $\{\mathcal{G}_t\}$ is of the form

$$\mathcal{G}_t = \sigma\big(\sigma_n^i, \chi_n^i | \sigma_n^i \le t, i = 1, \dots, K\big), \qquad t \ge 0, \tag{2.9}$$

where $(\sigma_n^i)_{n\in\mathbb{N}}$ is an increasing sequence of $\{\mathcal{F}_t^i\}$-stopping times and each χ_n^i is an $\mathcal{F}^i_{\sigma_n^i}$-measurable statistic that takes values in a *finite* set. In other words, a decentralized estimator must rely on quantized versions of the sensor observations, which may be transmitted to the fusion center at stopping times of the local sensor filtrations.

If the fusion center learns the complete sensor observations at any time t, then it can construct $\{\mathcal{F}_t\}$-adapted estimators, which we will call *centralized*. Assuming that for every $\lambda \in \mathbb{R}$,

$$\text{(A2)} \qquad \mathsf{P}_\lambda(A_t > 0) = 1 \qquad \forall t > 0,$$

the centralized, fixed-horizon MLE of λ at some time $t > 0$ is

$$\hat{\lambda}_t := \frac{B_t}{A_t}, \tag{2.10}$$

that is, the maximizer of the corresponding log-likelihood function,

$$\ell_t(\lambda) := \log \frac{\mathrm{d}\mathsf{P}_\lambda}{\mathrm{d}\mathsf{P}_0}\bigg|_{\mathcal{F}_t} = \lambda B_t - \frac{\lambda^2}{2} A_t. \tag{2.11}$$

From (2.11) we also obtain the corresponding score process and (observed) Fisher information, that is,

$$M_t := \frac{\mathrm{d}\ell_t(\lambda)}{\mathrm{d}\lambda} = B_t - \lambda A_t, \qquad -\frac{\mathrm{d}^2\ell_t(\lambda)}{\mathrm{d}\lambda^2} = A_t, \qquad t \geq 0, \tag{2.12}$$

and, consequently, we have

$$\hat{\lambda}_t = \lambda + \frac{M_t}{A_t}, \qquad t > 0. \tag{2.13}$$

Moreover, from (2.4), (2.5) and (2.8) it follows that $M \in \mathcal{M}_\lambda$, since

$$M_t = \sum_{i=1}^{K} \int_0^t X_s^i \,\mathrm{d}N_s^i, \qquad t \geq 0. \tag{2.14}$$

Since $\langle M, M\rangle = \langle B, B\rangle = A$, if we also assume that for every $\lambda \in \mathbb{R}$

$$\text{(A3)} \qquad \mathsf{P}_\lambda\Big(\lim_{t\to\infty} A_t = \infty\Big) = 1,$$

then there exists a P_λ-Brownian motion W (see [11], page 174) so that

$$\mathsf{P}_\lambda(M_t = W_{A_t}, t \geq 0) = 1. \tag{2.15}$$

This representation has some important consequences, which we state in the following lemma.

LEMMA 2.1. (a) *If $(t_\gamma)_{\gamma>0}$ is an increasing family of (possibly random) times so that $t_\gamma \to \infty$ P_λ-a.s., then $\hat{\lambda}_{t_\gamma} \to \lambda$ P_λ-a.s. as $\gamma \to \infty$.*

(b) *If $T_1 \leq T_2$ are $\{\mathcal{F}_t\}$-stopping times so that $\mathsf{E}_\lambda[A_{T_2}] < \infty$, then*

$$\mathsf{E}_\lambda[M_{T_1}] = \mathsf{E}_\lambda[M_{T_2}] = 0, \tag{2.16}$$

$$\mathsf{E}_\lambda\big[(M_{T_2} - M_{T_1})^2\big] = \mathsf{E}_\lambda[A_{T_2} - A_{T_1}]. \tag{2.17}$$

(c) *If* $\{A_t\}$ *is deterministic, then*

$$\sqrt{A_t}(\hat{\lambda}_t - \lambda) \sim \mathcal{N}(0, 1) \qquad \forall t > 0. \tag{2.18}$$

PROOF. Part (a) is a consequence of (2.13), (2.15) and the strong law of large numbers for the Brownian motion. Part (b) follows from a localization argument, optional sampling theorem and Doob's maximal inequality. Finally, when $\{A_t\}$ is deterministic, from (2.15) it follows that $M_t \sim \mathcal{N}(0, A_t)$ for every $t > 0$. From this observation and (2.13) we obtain (2.18). □

In the following lemma we state a version of the Cramer–Rao–Wolfowitz inequality.

LEMMA 2.2. *If* T *is an* $\{\mathcal{F}_t\}$*-stopping time and* ϕ *is an* $\mathcal{F}_T$*-measurable statistic so that* $0 < \mathsf{E}_\lambda[A_T] < \infty$ *and* $\mathsf{E}_\lambda[\phi] = \lambda$, $\mathsf{V}_\lambda[\phi] < \infty$ *for every* $\lambda \in \mathbb{R}$, *then*

$$\mathsf{V}_\lambda[\phi] \geq \frac{1}{\mathsf{E}_\lambda[A_T]}.$$

PROOF. From (2.16) and (2.17) and the Cauchy–Schwarz inequality we have

$$\mathsf{E}_\lambda[\phi M_T] = \mathsf{E}_\lambda\big[(\phi - \lambda) M_T\big] \leq \sqrt{\mathsf{E}_\lambda\big[(\phi - \lambda)^2\big] \mathsf{E}_\lambda\big[(M_T)^2\big]} = \sqrt{\mathsf{V}_\lambda[\phi] \mathsf{E}_\lambda[A_T]}.$$

Thus, it suffices to show that $\mathsf{E}_\lambda[\phi M_T] = 1$. Indeed, changing the measure $\mathsf{P}_\lambda \mapsto \mathsf{P}_0$ and differentiating both sides in $\mathsf{E}_\lambda[\phi] = \lambda$ with respect to λ,

$$1 = \frac{\mathrm{d}}{\mathrm{d}\lambda} \mathsf{E}_0\big[e^{\lambda B_T - (\lambda^2/2) A_T} \phi\big] = \mathsf{E}_0\big[e^{\lambda B_T - (\lambda^2/2) A_T} M_T \phi\big] = \mathsf{E}_\lambda[M_T \phi].$$

The second equality follows from interchanging derivative and expectation, which is possible due to the (quadratic) form of the log-likelihood function (2.11) (see, e.g., [12], page 54). □

Lemma 2.2 and (2.18) imply that when A_t is deterministic, $\hat{\lambda}_t$ is an optimal estimator of λ, in the sense that it has the smallest possible variance among $\mathcal{F}_t$-measurable, unbiased estimators (for any fixed $t > 0$). In order to obtain such an exact optimality property when $\{A_t\}$ is random, we consider the following sequential version of the centralized MLE:

$$\mathcal{S}_\gamma := \inf\{t \geq 0 : A_t \geq \gamma\}, \qquad \hat{\lambda}_{\mathcal{S}_\gamma} = \left(\frac{B}{A}\right)_{\mathcal{S}_\gamma}, \qquad \gamma > 0. \tag{2.19}$$

LEMMA 2.3. *For every* $\gamma > 0$,

$$\mathsf{P}_\lambda(\mathcal{S}_\gamma < \infty) = 1, \tag{2.20}$$

$$\sqrt{\gamma}(\hat{\lambda}_{\mathcal{S}_\gamma} - \lambda) \sim \mathcal{N}(0, 1). \tag{2.21}$$

Moreover, $\mathsf{P}_\lambda(\hat{\lambda}_{\mathcal{S}_\gamma} \to \lambda) = 1$ *as* $\gamma \to \infty$.

PROOF. Assumption (A3) implies (2.20). Since A has continuous paths, $A_{\mathcal{S}_\gamma} = \gamma$. Thus, from (2.15) we have $M_{\mathcal{S}_\gamma} \sim \mathcal{N}(0,\gamma)$ and, consequently, from (2.18) we obtain (2.21). Finally, the strong consistency of $\hat{\lambda}_{\mathcal{S}_\gamma}$ as $\gamma \to \infty$ is implied by Lemma 2.1(a). □

From Lemmas 2.2 and 2.3 it follows that, for any given $\gamma > 0$, $\hat{\lambda}_{\mathcal{S}_\gamma}$ is an optimal estimator of λ, in the sense that it has the smallest possible variance among unbiased, $\{\mathcal{F}_t\}$-adapted estimators (T_γ, ϕ_γ) for which $\mathsf{E}_\lambda[A_{T_\gamma}] \le \gamma$.

Therefore, there is always a centralized estimator of λ that is unbiased, normally distributed and optimal in a *nonasymptotic* sense. A decentralized estimator cannot have such a strong optimality property, as it relies on less information. However, we will say that a (decentralized) estimator is *asymptotically optimal*, if it has the same distribution as the corresponding optimal centralized estimator when an asymptotically large horizon of observations is available. More specifically,

(a) when $\{A_t\}$ is deterministic, a fixed-horizon, $\{\mathcal{G}_t\}$-adapted estimator $(\phi_t)_{t>0}$ will be *asymptotically optimal* if

$$\sqrt{A_t}(\phi_t - \lambda) \to \mathcal{N}(0,1) \qquad \text{as } t \to \infty,$$

(b) when $\{A_t\}$ is random, a sequential, $\{\mathcal{G}_t\}$-adapted estimator $(T_\gamma, \phi_\gamma)_{\gamma>0}$ will be *asymptotically optimal* if

$$\limsup_{\gamma\to\infty}\big(\mathsf{E}_\lambda[A_{T_\gamma}] - \gamma\big) \le 0 \quad \text{and} \quad \sqrt{\gamma}(\phi_\gamma - \lambda) \to \mathcal{N}(0,1) \qquad \text{as } \gamma \to \infty.$$

2.3. *Notation*. We close this section with some notation that will be useful in the construction and analysis of the proposed estimating scheme. Thus, for every $1 \le i \le K$ we define the statistic

$$B_t^i := \int_0^t X_s^i \, \mathrm{d}Y_s^i, \qquad t \ge 0, \tag{2.22}$$

and for any $1 \le i, j \le K$ we denote by A^{ij} the quadratic covariation of B^i and B^j and by A^i the quadratic variation of B^i, that is,

$$A_t^{ij} := \langle B^i, B^j\rangle_t = \int_0^t X_s^i X_s^j \, \mathrm{d}\langle Y^i, Y^j\rangle_s, \qquad t \ge 0, \tag{2.23}$$

$$A_t^{i} := \langle B^i, B^i\rangle_t = \int_0^t (X_s^i)^2 \, \mathrm{d}\langle Y^i, Y^i\rangle_s, \qquad t \ge 0. \tag{2.24}$$

Then, recalling the definitions of B and A in (2.4) and (2.5), we have

$$B = \sum_{i=1}^K B^i, \qquad A = \sum_{i=1}^K A^i + \sum_{1 \le i \ne j \le K} A^{ij}. \tag{2.25}$$

Moreover, we define the set

$$\mathcal{D} := \{(i, j) | 1 \le i \ne j \le K \text{ and } A^{ij} \text{ is random}\} \tag{2.26}$$

and we have the following representation for A:

$$A = \sum_{i=1}^{K} A^i + \sum_{(i,j)\in\mathcal{D}} A^{ij} + \sum_{(i,j)\notin\mathcal{D}} A^{ij}. \tag{2.27}$$

3. A decentralized estimating scheme. In this section we construct and analyze the proposed decentralized estimator. More specifically, we first define the communication scheme at the sensors and then introduce the statistics and estimators that will be used by the fusion center. As in the centralized setup, we distinguish two cases and consider a fixed-horizon estimator when $\{A_t\}$ is deterministic and a sequential estimator when $\{A_t\}$ is random. In each case, we analyze the asymptotic behavior of the resulting estimator as the horizon of observations goes to infinity *and* the rate of communication goes to zero, assuming that conditions (A1), (A2), (A3) are satisfied.

The main idea in the suggested communication scheme is that each sensor should inform the fusion center about the sufficient statistics for λ that it observes locally. However, instead of communicating at deterministic times, its communication times should be triggered by its local observations. In other words, each sensor i should inform the fusion center about the evolution of the $\{\mathcal{F}^i_t\}$-adapted, sufficient statistics for λ at a sequence of $\{\mathcal{F}^i_t\}$-stopping times.

When A is deterministic, $B^1, \dots, B^K$ are the only sufficient statistics for λ and it is clear that each B^i is $\{\mathcal{F}^i_t\}$-adapted, thus observable at sensor i.

When A is random, there are additional sufficient statistics, the *random* processes of the form A^i or A^{ij} (when A^i or A^{ij} is deterministic, it is completely known to the fusion center at any time t). If A^i is random, it is clear that it is $\{\mathcal{F}^i_t\}$-adapted, since $\langle B^i, B^i\rangle = A^i$. On the other hand, if A^{ij} (with $i \ne j$) is random, it is not locally observed either at sensor i or at sensor j, thus, the fusion center cannot be informed about its evolution (since there is no communication between sensors).

3.1. *Communication scheme and fusion center statistics*. Based on the previous discussion, we suggest that each sensor i communicate with the fusion center at the times

$$\tau^{i,B}_n := \inf\{t \ge \tau^{i,B}_{n-1} : B^i_t - B^i_{\tau^{i,B}_{n-1}} \notin (-\underline{\Delta}^i, \overline{\Delta}^i)\}, \qquad n \in \mathbb{N}, \tag{3.1}$$

and, *if A and A^i are random*, also at the times

$$\tau^{i,A}_n := \inf\{t \ge \tau^{i,A}_{n-1} : A^i_t - A^i_{\tau^{i,A}_{n-1}} \ge c^i\}, \qquad n \in \mathbb{N}, \tag{3.2}$$

where $\tau_0^{i,A} = \tau_0^{i,B} := 0$ and $c^i, \overline{\Delta}^i, \underline{\Delta}^i > 0$ are arbitrary, constant thresholds, chosen by the designer of the scheme, known both at sensor i and the fusion center. If either A^i or A is deterministic, sensor i does not communicate at the times $(\tau_n^{i,A})$ and we set $\tau_n^{i,A} = \infty$ for every $n \geq 1$.

At $\tau_n^{i,B}$, sensor i transmits to the fusion center with one bit the outcome of the Bernoulli random variable

$$z_n^i := \begin{cases} 1, & \text{if } B^i_{\tau_n^{i,B}} - B^i_{\tau_{n-1}^{i,B}} \geq \overline{\Delta}^i, \\ 0, & \text{if } B^i_{\tau_n^{i,B}} - B^i_{\tau_{n-1}^{i,B}} \leq -\underline{\Delta}^i, \end{cases} \tag{3.3}$$

whereas at $\tau_n^{i,A}$, if needed, it informs the fusion center with one bit that A^i has increased by c^i since $\tau_{n-1}^{i,A}$. Therefore, the induced filtration at the fusion center is

$$\tilde{\mathcal{F}}_t := \sigma\big(\tau_n^{i,A}, \tau_n^{i,B}, z_n^i | \tau_n^{i,A} \leq t, \tau_n^{i,B} \leq t, i = 1, \dots, K\big), \qquad t \geq 0, \tag{3.4}$$

which means that the fusion center can compute any $\{\tilde{\mathcal{F}}_t\}$-adapted statistic.

For every $1 \leq i \leq K$ we define

$$\tilde{A}_t^i := nc^i, \qquad \tau_n^{i,A} \leq t < \tau_{n+1}^{i,A}, \qquad n \in \mathbb{N} \cup \{0\}, \tag{3.5}$$

$$\tilde{B}_t^i := \sum_{j=1}^n [\overline{\Delta}^i z_j^i - \underline{\Delta}^i (1 - z_j^i)], \qquad \tau_n^{i,B} \leq t < \tau_{n+1}^{i,B}, \qquad n \geq 1, \tag{3.6}$$

where $\tilde{B}_t^i := 0$ for $t < \tau_1^{i,B}$, with the understanding that $\tilde{A}^i := A^i$ when A^i is deterministic. Moreover, motivated by (2.25)–(2.27), we define

$$\tilde{B} := \sum_{i=1}^K \tilde{B}^i, \tag{3.7}$$

$$\begin{aligned} \tilde{A} &:= \sum_{i=1}^K \tilde{A}^i + \sum_{i=1}^K d_i \tilde{A}^i + \sum_{(i,j) \notin \mathcal{D}} A^{ij} \\ &= \sum_{i=1}^K (1 + d_i) \tilde{A}^i + \sum_{(i,j) \notin \mathcal{D}} A^{ij}, \end{aligned} \tag{3.8}$$

where d_i is the number of random terms of the form A^{ij}, that is,

$$d_i := \#\{j | 1 \leq i \neq j \leq K \text{ and } A^{ij} \text{ is random}\}. \tag{3.9}$$

Again, we set $\tilde{A} := A$ when A is deterministic. Finally, we define the following quantities:

$$\Delta := \sum_{i=1}^K \max\{\overline{\Delta}^i, \underline{\Delta}^i\}, \qquad c := \sum_{i=1}^K (1 + d_i) c^i, \tag{3.10}$$

which will play an important role in the asymptotic analysis of the proposed estimating scheme.

LEMMA 3.1. *For every* $1 \le i \le K$ *and* $t, c, \Delta > 0$,

$$0 \le A_t^i - \tilde{A}_t^i \le c^i, \qquad |B_t^i - \tilde{B}_t^i| \le \max\{\overline{\Delta}^i, \underline{\Delta}^i\}, \tag{3.11}$$

$$A_t - \tilde{A}_t \le c, \qquad |B_t - \tilde{B}_t| \le \Delta. \tag{3.12}$$

PROOF. If A^i, A are deterministic, then $\tilde{A}^i := A^i$, $\tilde{A} := A$ and the corresponding inequalities hold trivially. Thus, without loss of generality, we assume that both A^i and A are random.

First of all, we observe that $\tilde{B}^i$ is exactly equal to B^i at $\tau_n^{i,B}$ and $\tilde{A}^i$ is exactly equal to A^i at $\tau_n^{i,A}$ for every $n \in \mathbb{N}$. Indeed, due to the path continuity of A^i and B^i, for every $n \in \mathbb{N}$ it is

$$\tilde{A}^i_{\tau_n^{i,A}} = nc^i = \sum_{j=1}^{n} [A^i_{\tau_j^{i,A}} - A^i_{\tau_{j-1}^{i,A}}] = A^i_{\tau_n^{i,A}},$$

$$\tilde{B}^i_{\tau_n^{i,B}} = \sum_{j=1}^{n} [\overline{\Delta}^i z_j^i - \underline{\Delta}^i (1 - z_j^i)] = \sum_{j=1}^{n} [B^i_{\tau_j^{i,B}} - B^i_{\tau_{j-1}^{i,B}}] = B^i_{\tau_n^{i,B}}.$$

Moreover, from the definition of the communication times $(\tau_n^{i,B})_n$, it is clear that $|B_t^i - \tilde{B}_t^i| < \max\{\overline{\Delta}^i, \underline{\Delta}^i\}$ for any time t between two jump times of $\tilde{B}^i$, which proves the second inequality in (3.11). Similarly, from the definition of $(\tau_n^{i,A})_n$ and the fact that A^i has increasing paths, it is clear that $0 < A_t^i - \tilde{A}_t^i < c^i$ for any time t between two jump times of $\tilde{A}^i$, which proves the first inequality in (3.11).

The second inequality in (3.12) follows directly from the second inequality in (3.11) and the definition of Δ. Finally, from the Kunita–Watanabe inequality (see [11], page 142) and the algebraic inequality $2\sqrt{|xy|} \le |x| + |y|$ we have

$$|A^{ij}| \le \sqrt{A^i A^j} \le \tfrac{1}{2}(A^i + A^j), \qquad 1 \le i \ne j \le K,$$

thus, from the definitions of $\mathcal{D}$ and d_i [recall (2.26) and (3.9)] we obtain

$$\sum_{(i,j)\in\mathcal{D}} A^{ij} \le \frac{1}{2} \sum_{(i,j)\in\mathcal{D}} (A^i + A^j) = \sum_{(i,j)\in\mathcal{D}} A^i = \sum_{i=1}^{K} d_i A^i.$$

From the representation of A in (2.27) and the latter inequality we have

$$\begin{aligned} A &\le \sum_{i=1}^{K} A^i + \sum_{i=1}^{K} d_i A^i + \sum_{(i,j)\notin\mathcal{D}} A^{ij} \\ &\le \sum_{i=1}^{K} (1 + d_i)(\tilde{A}^i + c^i) + \sum_{(i,j)\notin\mathcal{D}} A^{ij} = \tilde{A} + c, \end{aligned}$$

where the second inequality is due to (3.11) and the equality follows from the definitions of $\tilde{A}$ and c in (3.8) and (3.10), respectively. □

3.2. *The proposed estimator.* The proposed communication scheme requires the transmission of only one bit whenever a sensor communicates with the fusion center. Thus, the overall communication activity in the network will be low as long as the communication rate of each sensor is low. Therefore, we should ideally design an $\{\tilde{\mathcal{F}}_t\}$-adapted estimator that is statistically efficient even under an asymptotically low communication rate as the horizon of observations goes to infinity. For this reason, we let $\Delta \to \infty$ and $c \to \infty$ as $t \to \infty$ (or $\gamma \to \infty$) and we determine the relative rates that guarantee consistency and asymptotic optimality.

When $\{A_t\}$ is *deterministic*, we suggest the following estimator of λ at some arbitrary, deterministic time $t > 0$:

$$\tilde{\lambda}_t := \frac{\tilde{B}_t}{A_t}. \tag{3.13}$$

In the following theorem, which is the first main result of this paper, we show that $\{\tilde{\lambda}_t\}$ is consistent and asymptotically optimal under an asymptotically low communication rate.

THEOREM 3.1. *If* $t, \Delta \to \infty$ *so that* $\Delta = o(A_t)$, *then* $\tilde{\lambda}_t$ *converges to* λ *almost surely and in mean square. If additionally* $\Delta = o(\sqrt{A_t})$, *then* $\tilde{\lambda}_t$ *is asymptotically optimal, that is,* $\sqrt{A_t}(\tilde{\lambda}_t - \lambda) \to \mathcal{N}(0, 1)$.

PROOF. Since $\hat{\lambda}_t$ converges to λ almost surely and in mean square as $t \to \infty$, in order to prove that $\tilde{\lambda}_t$ is consistent, it suffices to show that $\mathsf{P}_\lambda(|\tilde{\lambda}_t - \hat{\lambda}_t| \to 0) = 1$ and $\mathsf{E}_\lambda[(\tilde{\lambda}_t - \hat{\lambda}_t)^2] \to 0$ as $t, \Delta \to \infty$ so that $\Delta = o(A_t)$.

Moreover, since $\sqrt{A_t}(\hat{\lambda}_t - \lambda) \sim \mathcal{N}(0, 1)$ for any $t > 0$, in order to establish the asymptotic optimality of $\tilde{\lambda}_t$, it suffices to show that $\sqrt{A_t}|\tilde{\lambda}_t - \hat{\lambda}_t|$ converges to 0 in probability as $t, \Delta \to \infty$ so that $\Delta = o(\sqrt{A_t})$.

Indeed, from the second inequality in (3.12) we have

$$|\tilde{\lambda}_t - \hat{\lambda}_t| = \left|\frac{\tilde{B}_t}{A_t} - \frac{B_t}{A_t}\right| = \frac{|\tilde{B}_t - B_t|}{A_t} \leq \frac{\Delta}{A_t}, \qquad t > 0,$$

which proves both claims. □

When $\{A_t\}$ is *random*, we suggest the following sequential, $\{\tilde{\mathcal{F}}_t\}$-adapted estimator of λ:

$$\tilde{\mathcal{S}}_\gamma := \inf\{t \geq 0 : \tilde{A}_t \geq \gamma - c\}, \qquad \tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} := \left(\frac{\tilde{B}}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma}, \qquad \gamma > c. \tag{3.14}$$

LEMMA 3.2. *For any γ, c such that $\gamma > c$,*

$$\mathsf{P}_\lambda(\tilde{\mathcal{S}}_\gamma \le \mathcal{S}_\gamma < \infty) = 1, \tag{3.15}$$

$$\mathsf{P}_\lambda(A_{\tilde{\mathcal{S}}_\gamma} \le \gamma) = 1, \tag{3.16}$$

$$\mathsf{E}_\lambda\big[(M_{\tilde{\mathcal{S}}_\gamma})^2\big] \le \gamma. \tag{3.17}$$

Moreover, if $c, \gamma \to \infty$ so that $c = o(\gamma)$, then

$$\limsup_{\gamma\to\infty}\big(\mathsf{E}_\lambda[A_{\tilde{\mathcal{S}}_\gamma}] - \gamma\big) \le 0. \tag{3.18}$$

PROOF. From the first inequality in (3.12) we have $\tilde{A} \ge A - c$, therefore,

$$\tilde{\mathcal{S}}_\gamma \le \inf\{t \ge 0 : A_t - c \ge \gamma - c\} = \mathcal{S}_\gamma. \tag{3.19}$$

From this inequality and (2.20) we obtain (3.15). Moreover, since A is the quadratic variation of B, it has continuous and increasing paths, thus, from (3.15) we obtain $\mathsf{P}_\lambda(A_{\tilde{\mathcal{S}}_\gamma} \le A_{\mathcal{S}_\gamma} = \gamma) = 1$. Finally, from (2.17) and (3.16) we obtain

$$\mathsf{E}_\lambda\big[(M_{\tilde{\mathcal{S}}_\gamma})^2\big] = \mathsf{E}_\lambda[A_{\tilde{\mathcal{S}}_\gamma}] \le \gamma,$$

which proves (3.17) and implies (3.18). □

In the following theorem we show that $\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma}$ is a consistent estimator of λ, even under an asymptotically low communication rate.

THEOREM 3.2. $\mathsf{P}_\lambda(\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} \to \lambda) = 1$ *and* $\mathsf{E}_\lambda[(\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \lambda)^2] \to 0$ *as* $\gamma, c, \Delta \to \infty$ *so that* $c, \Delta = o(\gamma)$.

PROOF. Recalling from (2.12) that $B = \lambda A + M$, we have P_λ-a.s.

$$\begin{aligned}
\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} &= \left(\frac{\tilde{B}}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} = \left(\frac{\tilde{B} - B}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} + \left(\frac{B}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} \\
&= \left(\frac{\tilde{B} - B}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} + \lambda\left(\frac{A}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} + \left(\frac{M}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma}
\end{aligned}$$

and, consequently,

$$\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \lambda = \left(\frac{\tilde{B} - B}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} + \lambda\left(\frac{A - \tilde{A}}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} + \left(\frac{M}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma}. \tag{3.20}$$

From the definition of $\tilde{\mathcal{S}}_\gamma$ it follows that $\tilde{A}_{\tilde{\mathcal{S}}_\gamma} \ge \gamma - c$, whereas from (3.12) we have $|\tilde{B} - B|_{\tilde{\mathcal{S}}_\gamma} \le \Delta$ and $(A - \tilde{A})_{\tilde{\mathcal{S}}_\gamma} \le c$. Therefore,

$$|\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \lambda| \le \frac{\Delta + |\lambda| c}{\gamma - c} + \frac{|M_{\tilde{\mathcal{S}}_\gamma}|}{\gamma - c}. \tag{3.21}$$

The first term in the right-hand side clearly goes to 0 as $c, \Delta, \gamma \to \infty$ so that $c, \Delta = o(\gamma)$. Moreover, from (2.15) and (3.16) we have P_λ-a.s.

$$\frac{|M_{\tilde{\mathcal{S}}_\gamma}|}{\gamma - c} = \frac{|W_{A_{\tilde{\mathcal{S}}_\gamma}}|}{A_{\tilde{\mathcal{S}}_\gamma}} \frac{A_{\tilde{\mathcal{S}}_\gamma}}{\gamma - c} \leq \frac{|W_{A_{\tilde{\mathcal{S}}_\gamma}}|}{A_{\tilde{\mathcal{S}}_\gamma}} \frac{\gamma}{\gamma - c}. \tag{3.22}$$

If $c, \gamma \to \infty$ so that $c = o(\gamma)$, $\mathsf{P}_\lambda(A_{\tilde{\mathcal{S}}_\gamma} \to \infty) = 1$, due to assumption (A3). Therefore, the strong law of large numbers implies that the right-hand side in (3.22) converges to 0 and, consequently, $\mathsf{P}_\lambda(\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} \to \lambda) = 1$ as $c, \gamma \to \infty$ so that $c = o(\gamma)$.

Moreover, if we square both sides in (3.21), apply the algebraic inequality $(x + y)^2 \leq 2(x^2 + y^2)$, take expectations and use (3.17), we obtain

$$\mathsf{E}_\lambda[(\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \lambda)^2] \leq 2\left(\frac{\Delta + |\lambda| c}{\gamma - c}\right)^2 + 2\frac{\gamma}{(\gamma - c)^2},$$

which implies that $\mathsf{E}_\lambda[(\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \lambda)^2] \to 0$ as $c, \Delta, \gamma \to \infty$ so that $c, \Delta = o(\gamma)$. □

The consistency of $\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma}$ was established without any additional conditions on the dynamics of the sensor processes. However, it is clear that the suggested estimator cannot be asymptotically efficient in such a general setup, since it does not have any access to sufficient statistics of the form A^{ij} with $(i, j) \in \mathcal{D}$.

Nevertheless, if every A^{ij} with $i \neq j$ is deterministic, then $\mathcal{D} = \varnothing$ and the fusion center has access to all sufficient statistics for λ. In this case, we can obtain an asymptotically sharp lower bound for $A_{\tilde{\mathcal{S}}_\gamma}$, the observed Fisher information that is utilized by the proposed estimator, which allows us to establish its asymptotic optimality even under an asymptotically low communication rate.

LEMMA 3.3. *If $\mathcal{D} = \varnothing$, then $\tilde{A}_t \leq A_t$ for every $t \geq 0$. Consequently, for every γ, c such that $\gamma > c$,*

$$\mathsf{P}_\lambda(A_{\tilde{\mathcal{S}}_\gamma} \geq \gamma - c) = 1, \tag{3.23}$$

$$\mathsf{E}_\lambda\big[(M_{\mathcal{S}_\gamma} - M_{\tilde{\mathcal{S}}_\gamma})^2\big] \leq c. \tag{3.24}$$

PROOF. If $\mathcal{D} = \varnothing$, then $d_i = 0$ for every $1 \leq i \leq K$, thus, from (2.25), (3.8) and the first inequality in (3.11) we obtain

$$\tilde{A} = \sum_{i=1}^{K} \tilde{A}^i + \sum_{1 \leq j \neq i \leq K} A^{ij} \leq \sum_{i=1}^{K} A^i + \sum_{1 \leq j \neq i \leq K} A^{ij} = A.$$

Then, from the definition of $\tilde{\mathcal{S}}_\gamma$ we have $\mathsf{P}_\lambda(A_{\tilde{\mathcal{S}}_\gamma} \geq \tilde{A}_{\tilde{\mathcal{S}}_\gamma} \geq \gamma - c) = 1$, which proves (3.23). Finally, from (2.17), (3.19) and (3.23) we obtain

$$\mathsf{E}_\lambda\big[(M_{\mathcal{S}_\gamma} - M_{\tilde{\mathcal{S}}_\gamma})^2\big] = \mathsf{E}_\lambda[A_{\mathcal{S}_\gamma} - A_{\tilde{\mathcal{S}}_\gamma}] = \mathsf{E}_\lambda[\gamma - A_{\tilde{\mathcal{S}}_\gamma}] \leq c,$$

which completes the proof. □

THEOREM 3.3. *If* $\mathcal{D} = \varnothing$, *then* $\sqrt{\gamma}(\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \lambda) \to \mathcal{N}(0,1)$ *as* $c, \Delta, \gamma \to \infty$ *so that* $c, \Delta = o(\sqrt{\gamma})$.

PROOF. Since $\sqrt{\gamma}(\hat{\lambda}_{\mathcal{S}_\gamma} - \lambda) \sim \mathcal{N}(0,1)$ for every $\gamma > 0$, it suffices to show that $\sqrt{\gamma}|\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \hat{\lambda}_{\mathcal{S}_\gamma}|$ converges to zero in probability as $\gamma, c, \Delta \to \infty$ so that $c, \Delta = o(\sqrt{\gamma})$. Indeed, from (2.13) and (3.20) we have P_λ-a.s.

$$\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \hat{\lambda}_{\mathcal{S}_\gamma} = \left(\frac{\tilde{B} - B}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} + \lambda\left(\frac{A - \tilde{A}}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} + \left(\frac{M}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} - \left(\frac{M}{A}\right)_{\mathcal{S}_\gamma}.$$

Since $\tilde{A}_{\tilde{\mathcal{S}}_\gamma} \geq \gamma - c$ and from (3.12) we have $|\tilde{B} - B|_{\tilde{\mathcal{S}}_\gamma} \leq \Delta$ and $(A - \tilde{A})_{\tilde{\mathcal{S}}_\gamma} \leq c$,

$$\sqrt{\gamma}|\tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma} - \hat{\lambda}_{\mathcal{S}_\gamma}| \leq \sqrt{\gamma}\frac{\Delta + |\lambda| c}{\gamma - c} + \sqrt{\gamma}\left|\left(\frac{M}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} - \left(\frac{M}{A}\right)_{\mathcal{S}_\gamma}\right|. \tag{3.25}$$

The first term in the right-hand side of (3.25) converges to 0 as $c, \Delta, \gamma \to \infty$ so that $c, \Delta = o(\sqrt{\gamma})$. Moreover, since $A_{\mathcal{S}_\gamma} = \gamma$ and $\tilde{A}_{\tilde{\mathcal{S}}_\gamma} \geq \gamma - c$,

$$\begin{aligned}\sqrt{\gamma}\left|\left(\frac{M}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} - \left(\frac{M}{A}\right)_{\mathcal{S}_\gamma}\right| &= \sqrt{\gamma}\left|\left(\frac{M}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} - \frac{M_{\tilde{\mathcal{S}}_\gamma}}{\gamma} + \frac{M_{\tilde{\mathcal{S}}_\gamma}}{\gamma} - \frac{M_{\mathcal{S}_\gamma}}{\gamma}\right| \\ &\leq \frac{1}{\sqrt{\gamma}}\left[|M_{\tilde{\mathcal{S}}_\gamma}|\frac{\gamma - \tilde{A}_{\tilde{\mathcal{S}}_\gamma}}{\tilde{A}_{\tilde{\mathcal{S}}_\gamma}} + |M_{\tilde{\mathcal{S}}_\gamma} - M_{\mathcal{S}_\gamma}|\right] \qquad (3.26) \\ &\leq \frac{1}{\sqrt{\gamma}}\left[|M_{\tilde{\mathcal{S}}_\gamma}|\frac{c}{\gamma - c} + |M_{\tilde{\mathcal{S}}_\gamma} - M_{\mathcal{S}_\gamma}|\right].\end{aligned}$$

From the Cauchy–Schwarz inequality, (3.17) and (3.24) we have

$$\mathsf{E}_\lambda[|M_{\tilde{\mathcal{S}}_\gamma}|] \leq \sqrt{\mathsf{E}_\lambda[M^2_{\tilde{\mathcal{S}}_\gamma}]} \leq \sqrt{\gamma},$$

$$\mathsf{E}_\lambda[|M_{\tilde{\mathcal{S}}_\gamma} - M_{\mathcal{S}_\gamma}|] \leq \sqrt{\mathsf{E}_\lambda[(M_{\tilde{\mathcal{S}}_\gamma} - M_{\mathcal{S}_\gamma})^2]} \leq \sqrt{c}.$$

Then, taking expectations in (3.26), we obtain

$$\sqrt{\gamma}\mathsf{E}_\lambda\left[\left|\left(\frac{M}{\tilde{A}}\right)_{\tilde{\mathcal{S}}_\gamma} - \left(\frac{M}{A}\right)_{\mathcal{S}_\gamma}\right|\right] \leq \frac{c}{\gamma - c} + \sqrt{\frac{c}{\gamma}}.$$

Therefore, the second term in the right-hand side of (3.25) converges to 0 in probability, due to Markov's inequality, as $c, \Delta, \gamma \to \infty$ so that $c = o(\gamma)$. This concludes the proof. □

COROLLARY 3.1. *If* $\mathcal{D} = \varnothing$, *then* $(\tilde{\mathcal{S}}_\gamma, \tilde{\lambda}_{\tilde{\mathcal{S}}_\gamma})$ *is asymptotically optimal as* $\gamma, c, \Delta \to \infty$ *so that* $c, \Delta = o(\sqrt{\gamma})$.

PROOF. This is a consequence of (3.18) and Theorem 3.3. □

3.3. *Remarks and examples.* For the implementation of the proposed estimator, the fusion center does not need to record the values of the communication times. It simply needs to keep track of $\tilde{B}^1, \dots, \tilde{B}^K$ and—if necessary—$\tilde{A}^1, \dots, \tilde{A}^K$, and update them whenever it receives a relevant message. Since these statistics are defined recursively, at most $2K$ values need to be stored at any given time.

Theorems 3.1, 3.2 and 3.3 remain valid if c and Δ are held fixed as $t \to \infty$ or $\gamma \to \infty$. Moreover, they remain valid if we use in the definitions of $\tau_n^{i,B}$ and $\tau_n^{i,A}$ time-varying, positive thresholds, $\overline{\Delta}_n^i$, $\underline{\Delta}_n^i$, c_n^i, so that

$$\overline{\Delta}_n^i \leq \overline{\Delta}^i, \qquad \underline{\Delta}_n^i \leq \underline{\Delta}^i, \qquad c_n^i \leq c^i \qquad \forall n \in \mathbb{N}.$$

Therefore, it may be possible to improve the performance of the proposed estimator by introducing linear or curved boundaries and optimizing over the additional parameters.

We close this section with some examples that illustrate our main results. Thus, let $\sigma_t := [\sigma_t^{ij}]$ be an $\{\mathcal{F}_t\}$-adapted, square matrix of size K, set $\alpha_t := \sigma_t \sigma_t'$, where σ_t' is the transpose of σ_t, and consider the following special case of model (2.8):

$$Y_t^i = \lambda \sum_{j=1}^K \int_0^t X_s^j \alpha_s^{ij} \, \mathrm{d}s + \sum_{j=1}^K \int_0^t \sigma_s^{ij} \, \mathrm{d}W_s^j, \qquad t \geq 0, 1 \leq i \leq K, \tag{3.27}$$

where $(W^1, \dots, W^K)$ is a K-dimensional P_λ-Brownian motion. The observed Fisher information $\{A_t\}$ then becomes

$$A_t = \sum_{i=1}^K \sum_{j=1}^K \int_0^t X_s^i X_s^j \alpha_s^{ij} \, \mathrm{d}s, \qquad t \geq 0. \tag{3.28}$$

In Theorem 3.1, we stated the asymptotic properties of the proposed estimator when A_t is deterministic. This assumption is clearly satisfied when there are real functions $b_i, \rho_{ij} : [0, \infty) \to \mathbb{R}$ so that $X_t^i = b_i(t)$ and $\alpha_t^{ij} = \rho_{ij}(t)$ for every $1 \leq i, j \leq K$, in which case

$$A_t = \sum_{i=1}^K \sum_{j=1}^K \int_0^t b_i(s) b_j(s) \rho_{ij}(s) \, \mathrm{d}s, \qquad t \geq 0, \tag{3.29}$$

and $(Y^1, \dots, Y^K)$ is a Gaussian process with independent increments. However, Theorem 3.1 also applies when $X_t^i = b_i(t)/Y_t^i$ and $\alpha_t^{ij} = \rho_{ij}(t) Y_t^i Y_t^j$, in which case A is still given by (3.29).

In Theorem 3.3, we proved that the proposed estimator is asymptotically optimal when A^{ij} is deterministic for every $i \neq j$. This condition is clearly satisfied when $\sigma^{ij} = 0$ for every $i \neq j$, in which case $Y^1, \dots, Y^K$ are independent,

$\alpha^{ii} = (\sigma^{ii})^2$ and (3.27), (3.28) become

$$Y_t^i = \lambda \int_0^t X_s^i \alpha_s^{ii}\, \mathrm{d}s + \int_0^t \sqrt{\alpha_s^{ii}}\, \mathrm{d}W_s^i, \qquad t \ge 0,$$

$$A_t = \sum_{i=1}^K \int_0^t (X_s^i)^2 \alpha_s^{ii}\, \mathrm{d}s, \qquad t \ge 0.$$

If, in particular, X^i is a nonzero constant and $\alpha^{ii} = Y^i$, then Y^i is a square-root diffusion, whereas if $X^i = Y^i$ and α^{ii} is a positive constant, then Y^i is an Ornstein–Uhlenbeck process.

4. The Brownian case. In this section we assume that $\langle Y^i, Y^j \rangle_t = 0$, $\langle Y^i, Y^i \rangle_t = t$ and $X_t^i = x_i$, where $x_i \neq 0$ is a known constant, for every $1 \le i \neq j \le K$ and $t \ge 0$.

Thus, $B_t^i = x_i Y_t^i$, $A_t^i = (x_i)^2 t$, $A_t = \sum_{i=1}^K A_t^i$ and (2.8) reduces to

$$Y_t^i = \lambda x_i t + N_t^i, \qquad t \ge 0, i = 1, \dots, K,$$

where $N^1, \dots, N^K$ are independent, standard Brownian motions under P_λ.

Since the filtrations $\{\mathcal{F}_t^1\}, \dots, \{\mathcal{F}_t^K\}$ are independent, for every $1 \le i \le K$ and $t > 0$ we have

$$\frac{\mathrm{d}\mathsf{P}_\lambda}{\mathrm{d}\mathsf{P}_0}\bigg|_{\mathcal{F}_t^i} = e^{\lambda B_t^i - (\lambda^2/2) A_t^i} = e^{\lambda B_t^i - (\lambda x_i)^2 t/2}. \tag{4.1}$$

We also assume, for simplicity, that $\overline{\Delta}^i = \underline{\Delta}^i = \Delta^i$ for every $1 \le i \le K$, thus, $\Delta = \sum_{i=1}^K \Delta^i$ and

$$\tau_n^{i,B} = \inf\{t \ge \tau_{n-1}^{i,B} : |B_t^i - B_{\tau_{n-1}^{i,B}}^i| \ge \Delta^i\}, \tag{4.2}$$

$$z_n^i = \begin{cases} 1, & \text{if } B_{\tau_n^{i,B}}^i - B_{\tau_{n-1}^{i,B}}^i \ge \Delta^i, \\ 0, & \text{if } B_{\tau_n^{i,B}}^i - B_{\tau_{n-1}^{i,B}}^i \le -\Delta^i. \end{cases} \tag{4.3}$$

We denote by δ_n^i the time between the arrival of the $(n-1)$th and the nth message from sensor i and by m_t^i the number of transmitted messages by sensor i up to time t, that is,

$$\delta_n^i := \tau_n^{i,B} - \tau_{n-1}^{i,B}, \qquad m_t^i := \max\{n \in \mathbb{N} : \tau_n^i \le t\}. \tag{4.4}$$

Since $\{A_t\}$ is deterministic, $\tau_n^{i,A} = \infty$ for every $1 \le i \le K$ and $n \in \mathbb{N}$ and the fusion center filtration becomes

$$\tilde{\mathcal{F}}_t = \sigma(\delta_n^i, z_n^i; n \le m_t^i, 1 \le i \le K), \qquad t \ge 0.$$

Moreover, $\tilde{A} := A$ and $\tilde{A}^i := A^i$ for every i, however, we now define the following $\{\tilde{\mathcal{F}}_t\}$-adapted statistics:

$$\check{A}_t^i := |x_i|^2 \sum_{j=1}^{m_t^i} \delta_j^i, \qquad \check{A}_t := \sum_{i=1}^{K} \check{A}_t^i, \qquad t \geq 0. \tag{4.5}$$

That is, $\check{A}_t$ is an approximation of A_t that relies only on the communication times $\{\tau_n^{i,B}; n \leq m_t^i, 1 \leq i \leq K\}$.

Since Brownian motion "restarts" at stopping times, each $(\delta_n^i, z_n^i)_{n \in \mathbb{N}}$ is a sequence of i.i.d. pairs, thus, each $(m_t^i)_{t \geq 0}$ is a renewal process. Moreover, it is possible to obtain a series representation for the joint density of the pair (δ_1^i, z_1^i) under P_λ,

$$\bar{p}_i(t;\lambda) := \frac{\mathsf{P}_\lambda(\delta_1^i \in \mathrm{d}t, z_1^i = 1)}{\mathrm{d}t}, \qquad \underline{p}_i(t;\lambda) := \frac{\mathsf{P}_\lambda(\delta_1^i \in \mathrm{d}t, z_1^i = 0)}{\mathrm{d}t}.$$

This representation is the content of the following lemma, for which we need to define the following functions:

$$g(t;x) := \sum_{n=-\infty}^{\infty} h\big(t;(4n+1)x\big), \qquad h(t;x) := \frac{x}{\sqrt{2\pi t^3}} e^{-x^2/2t}, \qquad t, x \geq 0.$$

LEMMA 4.1. *For every* $1 \leq i \leq K$ *and* $t > 0$,

$$\bar{p}_i(t;\lambda) = e^{\lambda \Delta^i - 0.5(\lambda x_i)^2 t} g\big(t; \Delta^i/|x_i|\big),$$
$$\underline{p}_i(t;\lambda) = e^{-\lambda \Delta^i - 0.5(\lambda x_i)^2 t} g\big(t; \Delta^i/|x_i|\big).$$

PROOF. From (4.2) and (4.4) we have

$$\delta_1^i = \inf\{t \geq 0 : |Y_t^i| \geq \Delta^i/|x_i|\}, \qquad n \in \mathbb{N}. \tag{4.6}$$

Since Y^i is a standard Brownian motion under P_0, it is well known (see, e.g., [11], page 99) that $\bar{p}_i(t;0) = \underline{p}_i(t;0) = g(t;\Delta^i/|x_i|)$. Then, changing the measure $\mathsf{P}_\lambda \mapsto \mathsf{P}_0$ (similarly, e.g., to [11], page 196), we obtain the desired result. $\square$

The following lemma describes some properties of the communication scheme that remain valid in the case of discrete sampling at the sensors, which we treat in Section 4.2. In order to lighten the notation, we denote by $\Theta(\Delta^i)$ a term that when divided by Δ^i is asymptotically bounded from above and below as $\Delta^i \to \infty$.

LEMMA 4.2. (a) *For any* $t, \Delta^i > 0$,

$$\mathsf{E}_\lambda\left[\sum_{j=1}^{m_t^i+1} \delta_j^i - t\right] \leq \frac{\mathsf{E}_\lambda[(\delta_1^i)^2]}{\mathsf{E}_\lambda[\delta_1^i]}, \tag{4.7}$$

$$\mathsf{E}_\lambda\left[t - \sum_{j=1}^{m_t^i} \delta_j^i\right] \le \frac{\mathsf{E}_\lambda[(\delta_1^i)^2]}{\mathsf{E}_\lambda[\delta_1^i]}. \tag{4.8}$$

(b) *As* $t, \Delta^i \to \infty$,

$$\mathsf{E}_\lambda[\delta_1^i] = \Theta(\Delta^i), \qquad \mathsf{V}_\lambda[\delta_1^i] = \Theta(\Delta^i), \tag{4.9}$$

$$0 \le \mathsf{E}_\lambda[A_t^i - \check{A}_t^i] \le \Theta(\Delta^i), \tag{4.10}$$

$$\mathsf{E}_\lambda[m_t^i] \le t/\Theta(\Delta^i) + 1/\Theta(\Delta^i). \tag{4.11}$$

PROOF. (a) Since $(\delta_n^i)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables, (4.7) follows from Theorem 1 in Lorden [14] and (4.8) from Lorden [14], page 526.

(b) Recall from (4.6) that δ_1^i is the first time a Brownian motion with drift λx_i exits the symmetric interval $(-\Delta^i/|x_i|, \Delta^i/|x_i|)$. Then, as $\Delta^i \to \infty$, from Wald's identity we have

$$\mathsf{E}_\lambda[\delta_1^i] = \frac{\Delta^i/|x_i|}{|\lambda x_i|}(1 + o(1)), \tag{4.12}$$

whereas from Martinsek [16] we have

$$\mathsf{V}_\lambda[\delta_1^i] = \frac{\Delta^i/|x_i|}{|\lambda x_i|^3}(1 + o(1)). \tag{4.13}$$

Then, from (4.12) and (4.13) we obtain (4.9), whereas from (4.5), (4.8) and (4.9) we obtain (4.10).

Finally, since $m_t^i + 1$ is a stopping time with respect to the filtration generated by the pairs $(\delta_n^i, z_n^i)_{n\in\mathbb{N}}$, from Wald's identity and (4.7) we have

$$\mathsf{E}_\lambda[m_t^i + 1]\mathsf{E}_\lambda[\delta_1^i] = \mathsf{E}_\lambda\left[\sum_{j=1}^{m_t^i+1} \delta_j^i\right] \le t + \frac{\mathsf{E}_\lambda[(\delta_1^i)^2]}{\mathsf{E}_\lambda[\delta_1^i]}$$

and, consequently,

$$\mathsf{E}_\lambda[m_t^i] \le \frac{t}{\mathsf{E}_\lambda[\delta_1^i]} + \frac{\mathsf{V}_\lambda[\delta_1^i]}{(\mathsf{E}_\lambda[\delta_1^i])^2}.$$

From this inequality and (4.9) we obtain (4.11), which completes the proof. □

4.1. *Likelihood-based estimation at the fusion center.* Let $\tilde{\mathcal{L}}_t(\lambda)$ and $\tilde{\ell}_t(\lambda)$ be the likelihood and the log-likelihood function of λ that correspond to $\tilde{\mathcal{F}}_t$, the accumulated information at the fusion center up to time t. The following proposition describes the structure of the corresponding score function.

PROPOSITION 4.1. *For any $t > 0$,*

$$\frac{\mathrm{d}\tilde{\ell}_t(\lambda)}{\mathrm{d}\lambda} = \left\{\sum_{i=1}^K \mathsf{E}_\lambda[B_t^i|m_t^i] - \lambda A_t\right\} + \{\tilde{B}_t - \lambda\check{A}_t\}. \tag{4.14}$$

PROOF. Suppose that $m_t^i = m_i$, that is, sensor i has transmitted m_i messages to the fusion center up to time t, where m_i is some nonnegative integer. Then, since all pairs $\{(z_n^i, \delta_n^i), n \in \mathbb{N}, 1 \le i \le K\}$ are independent, the fusion likelihood function has the following form:

$$\tilde{\mathcal{L}}_t(\lambda) := \prod_{i=1}^K \mathsf{P}_\lambda(m_t^i = m_i)\left(\prod_{n=1}^{m_i} \bar{p}_i(\delta_n^i;\lambda)^{z_n^i} \cdot \underline{p}_i(\delta_n^i;\lambda)^{1-z_n^i}\right)^{\mathbb{1}_{\{m_i>0\}}}.$$

Due to Lemma 4.1, the corresponding log-likelihood function becomes

$$\begin{aligned}\tilde{\ell}_t(\lambda) &= \sum_{i=1}^K \log \mathsf{P}_\lambda(m_t^i = m_i)\\ &\quad + \sum_{i=1}^K \mathbb{1}_{\{m_i>0\}} \sum_{n=1}^{m_i}\left[\lambda\Delta^i - \frac{(\lambda x_i)^2\delta_n^i}{2} + \log g(\delta_n^i;\Delta^i/|x_i|)\right] z_n^i\\ &\quad + \sum_{i=1}^K \mathbb{1}_{\{m_i>0\}} \sum_{n=1}^{m_i}\left[-\lambda\Delta^i - \frac{(\lambda x_i)^2\delta_n^i}{2} + \log g(\delta_n^i;\Delta^i/|x_i|)\right](1 - z_n^i).\end{aligned}$$

Then, recalling the definition of $\tilde{B}$ in (3.6)–(3.7) and of $\check{A}$ in (4.5),

$$\frac{\mathrm{d}\tilde{\ell}_t(\lambda)}{\mathrm{d}\lambda} = \sum_{i=1}^K \frac{\mathrm{d}}{\mathrm{d}\lambda}\big(\log \mathsf{P}_\lambda(m_t^i = m_i)\big) + \tilde{B}_t - \lambda\check{A}_t.$$

Since $\{m_t^i = m_i\} \in \mathcal{F}_t^i$, changing the measure $\mathsf{P}_\lambda \mapsto \mathsf{P}_0$, we have

$$\mathsf{P}_\lambda(m_t^i = m_i) = \mathsf{E}_0\big[e^{\lambda B_t^i - \lambda^2 A_t^i/2}\mathbb{1}_{\{m_t^i = m_i\}}\big]$$

and, consequently,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\big(\log \mathsf{P}_\lambda(m_t^i = m_i)\big) &= \frac{\mathsf{E}_0[e^{\lambda B_t^i - \lambda^2 A_t^i/2}(B_t^i - \lambda A_t^i)\mathbb{1}_{\{m_t^i = m_i\}}]}{\mathsf{P}_\lambda(m_t^i = m_i)}\\ &= \frac{\mathsf{E}_\lambda[B_t^i \mathbb{1}_{\{m_t^i = m_i\}}] - \lambda A_t^i \mathsf{P}_\lambda(m_t^i = m_i)}{\mathsf{P}_\lambda(m_t^i = m_i)}\\ &= \mathsf{E}_\lambda[B_t^i | m_t^i = m_i] - \lambda A_t^i,\end{aligned}$$

which implies (4.14). □

Note that the second term in (4.14) reflects the information from the communication times and the transmitted messages, whereas the first term reflects the information between transmissions.

At time t, the fusion center should ideally estimate λ with the fusion center MLE, that is, the root of the score function (4.14). However, since $\mathsf{E}_\lambda[B_t^i | m_t^i]$ does not admit a simple, closed-form expression as a function of λ, we can only approximate this conditional expectation and obtain an *approximate* fusion center MLE.

If we replace each $\mathsf{E}_\lambda[B_t^i | m_t^i]$ with the corresponding unconditional expectation, $\mathsf{E}_\lambda[B_t^i] = \lambda A_t^i$, the first term in (4.14) vanishes and we obtain the following estimator:

$$\check{\lambda}_t := \frac{\tilde{B}_t}{\check{A}_t}, \qquad t \geq \min_{1 \leq i \leq K} \tau_1^i. \tag{4.15}$$

On the other hand, if we approximate $\mathsf{E}_\lambda[B_t^i | m_t^i]$ with $\lambda \check{A}_t^i$, we recover the estimator $\{\tilde{\lambda}_t\}$ that was defined in (3.13) and whose asymptotic properties were established in Theorem 3.1. In the following proposition we show that, in the special Brownian case that we consider in this section, $\check{\lambda}_t$ has similar asymptotic behavior as $\tilde{\lambda}_t$.

PROPOSITION 4.2. *If $t, \Delta \to \infty$ so that $\Delta = o(t)$, then $\check{\lambda}_t$ converges to λ in probability. If additionally $\Delta = o(\sqrt{t})$, then $\sqrt{A_t}(\check{\lambda}_t - \lambda) \to \mathcal{N}(0, 1)$, that is, $\check{\lambda}_t$ is an asymptotically optimal estimator of λ.*

PROOF. From the definition of $\tilde{\lambda}_t$ in (3.13) and $\check{\lambda}_t$ in (4.15) we have

$$\check{\lambda}_t - \tilde{\lambda}_t = \frac{\tilde{B}_t}{\check{A}_t} - \frac{\tilde{B}_t}{A_t} = \frac{A_t}{\check{A}_t} \frac{A_t - \check{A}_t}{A_t} \tilde{\lambda}_t, \qquad t \geq 0. \tag{4.16}$$

From (4.10) it follows that

$$0 \leq \frac{\mathsf{E}_\lambda[A_t - \check{A}_t]}{A_t} = \frac{1}{A_t} \sum_{i=1}^{K} \mathsf{E}_\lambda[A_t^i - \check{A}_t^i] \leq \sum_{i=1}^{K} \frac{\Theta(\Delta^i)}{A_t} = \frac{\Theta(\Delta)}{A_t}. \tag{4.17}$$

Therefore, Markov's inequality implies that $(A_t - \check{A}_t)/A_t$ converges to 0 and $A_t/\check{A}_t$ converges to 1 in probability as $t, \Delta \to \infty$ so that $\Delta = o(t)$, since A_t is a linear function of t. Moreover, from Theorem 3.1 we know that $\tilde{\lambda}_t$ converges to λ in probability if $\Delta = o(t)$. Thus, we conclude that $\check{\lambda}_t$ also converges to λ in probability as $t, \Delta \to \infty$ so that $\Delta = o(t)$.

In order to prove that $\check{\lambda}_t$ is asymptotically optimal, it suffices to show that $\sqrt{A_t}|\check{\lambda}_t - \tilde{\lambda}_t|$ converges to 0 in probability as $t, \Delta \to \infty$ so that $\Delta = o(\sqrt{t})$, which also follows from (4.16) and (4.17). □

4.2. *The case of discrete sampling*. We now assume that each sensor observes its underlying process only at a sequence of discrete and equidistant times $\{nh, n \in \mathbb{N}\}$, where $h > 0$ is a common sampling period. Thus, in what follows, $t = h, 2h, \dots$. The goal is to examine the effect of discrete sampling on the proposed estimating scheme.

First of all, we observe that the centralized estimator,

$$\hat{\lambda}_t = \frac{B_t}{A_t} = \frac{\sum_{i=1}^K x_i Y_t^i}{\sum_{i=1}^K (x_i)^2 t} \tag{4.18}$$

is not affected by the discrete sampling of the underlying processes and (2.18) remains valid, that is, $\sqrt{A_t}(\hat{\lambda}_t - \lambda) \sim \mathcal{N}(0, 1)$ for every $t = h, 2h, \dots$.

Moreover, the pairs $(\delta_n^i, z_n^i)_{n \in \mathbb{N}}$ remain i.i.d. and Lemma 4.2 still holds. On the other hand, Lemma 4.1 is no longer valid and there is not an explicit formula for the density of the pair (δ_1^i, z_1^i). However, the main difference in the case of discrete sampling is that at any time $\tau_n^{i,B}$ the fusion center learns whether B^i increased or decreased by at least Δ^i since $\tau_{n-1}^{i,B}$, but does not learn by how much exactly. In other words, the fusion center does not learn the size of the realized overshoots,

$$\eta_n^i := \left(B^i_{\tau_n^{i,B}} - B^i_{\tau_{n-1}^{i,B}} - \Delta^i\right)^+ + \left(B^i_{\tau_n^{i,B}} - B^i_{\tau_{n-1}^{i,B}} + \Delta^i\right)^-, \qquad n \in \mathbb{N}. \tag{4.19}$$

As a result, the statistic $\tilde{B}^i$, defined in (3.6), is no longer equal to B^i at the communication times $(\tau_n^{i,B})_{n \in \mathbb{N}}$ and the distance $|B_t^i - \tilde{B}_t^i|$ is no longer bounded by Δ^i. Therefore, Theorem 3.1, which establishes the consistency and asymptotic optimality of the proposed estimator, $\tilde{\lambda}_t = \tilde{B}_t / A_t$, under the assumption of continuous-time sensor observations may not hold when the sensors observe their underlying processes at discrete times.

Our goal is to determine under what conditions the consistency and asymptotic optimality of $\tilde{\lambda}_t$ are preserved in the context of discrete sampling at the sensors. In order to do so, we need to estimate the inflicted performance loss due to the unobserved overshoots. The following lemma is very useful in this direction.

LEMMA 4.3. *For every* $1 \le i \le K$,

$$\left|B_t^i - \tilde{B}_t^i\right| \le \Delta^i + \sum_{j=1}^{m_t^i} \eta_j^i, \qquad t \ge 0, \tag{4.20}$$

and the overshoots $(\eta_n^i)_{n \in \mathbb{N}}$ *are i.i.d. with*

$$\sup_{\Delta^i > 0} \mathsf{E}_\lambda[\eta_1^i] = \mathcal{O}(\sqrt[3]{h}). \tag{4.21}$$

PROOF. For every $t \geq 0$ we have

$$\begin{aligned} B_t^i - \tilde{B}_t^i &= B_t^i - B^i_{\tau^{i,B}_{m_t^i}} + \sum_{j=1}^{m_t^i} \big(B^i_{\tau_j^{i,B}} - B^i_{\tau_{j-1}^{i,B}}\big) - \tilde{B}_t^i \\ &= B_t^i - B^i_{\tau^{i,B}_{m_t^i}} + \sum_{j=1}^{m_t^i} \big[\big(B^i_{\tau_j^{i,B}} - B^i_{\tau_{j-1}^{i,B}}\big) - \big[\Delta^i z_j^i - \Delta^i(1 - z_j^i)\big]\big], \end{aligned}$$

which implies (4.20). It is obvious that the overshoots $(\eta_n^i)_{n \in \mathbb{N}}$ are i.i.d. In order to prove (4.21), we write $\delta_1^i = \min\{\underline{\delta}_1^i, \overline{\delta}_1^i\}$, where

$$\underline{\delta}_1^i := \inf\{nh : B_{nh}^i \leq -\Delta^i\}, \qquad \overline{\delta}_1^i := \inf\{nh : B_{nh}^i \geq \Delta^i\}.$$

Then, from Theorem 3 of Lorden [14] it follows that for *any* $r \geq 1$,

$$\begin{aligned} \sup_{\Delta^i > 0} \mathsf{E}_\lambda[\eta_1^i] &\leq \max\{\mathsf{E}_\lambda[B^i_{\overline{\delta}_1^i} - \Delta^i], -\mathsf{E}_\lambda[B^i_{\underline{\delta}_1^i} + \Delta^i]\} \\ &\leq \sqrt[r]{\frac{r+2}{r+1} \frac{\mathsf{E}_\lambda[|B_h^i|^{r+1}]}{|\mathsf{E}_\lambda[B_h^i]|}}. \end{aligned} \tag{4.22}$$

Since $Y_h^i \sim \mathcal{N}(\lambda x_i h, h)$ under P_λ and $B_h^i = x_i Y_h^i$,

$$\begin{aligned} \mathsf{E}_\lambda[B_h^i] &= \lambda (x_i)^2 h, \\ \mathsf{E}_\lambda[(B_h^i)^4] &= (x_i)^4[(\lambda x_i h)^4 + 6(\lambda x_i h)^2 h + 3h^2] \\ &= 3(x_i)^4 h^2(1 + o(1)) \qquad \text{as } h \to 0. \end{aligned}$$

Setting $r = 3$ in (4.22) completes the proof. □

In the following theorem we show that $\tilde{\lambda}_t$ remains consistent as $t \to \infty$ for any given, fixed sampling period, $h > 0$, as long as the communication rate of every sensor is asymptotically low.

THEOREM 4.1. *If* $t, \Delta^i \to \infty$ *so that* $\Delta^i = o(t)$ *for every* $1 \leq i \leq K$, *then* $\mathsf{E}_\lambda[|\tilde{\lambda}_t - \lambda|] \to 0$.

PROOF. Since $\mathsf{E}_\lambda[|\hat{\lambda}_t - \lambda|] \to 0$, it suffices to show that $\mathsf{E}_\lambda[|\tilde{\lambda}_t - \hat{\lambda}_t|] \to 0$. Indeed, from the definition of the two estimators and (4.20) we have

$$|\tilde{\lambda}_t - \hat{\lambda}_t| \leq \frac{1}{A_t} \sum_{i=1}^K |\tilde{B}_t^i - B_t^i| \leq \frac{\Delta}{A_t} + \frac{1}{A_t} \sum_{i=1}^K \sum_{j=1}^{m_t^i + 1} \eta_j^i. \tag{4.23}$$

Since $m_t^i + 1$ is a stopping time with respect to the filtration generated by $(\delta_n^i, z_n^i, \eta_n^i)_{n\in\mathbb{N}}$, from Wald's identity we obtain

$$\mathsf{E}_\lambda\left[\sum_{j=1}^{m_t^i+1} \eta_j^i\right] = \mathsf{E}_\lambda[\eta_1^i]\mathsf{E}_\lambda[m_t^i + 1]. \tag{4.24}$$

Taking expectations in (4.23) and applying (4.24), we obtain

$$\mathsf{E}_\lambda[|\tilde{\lambda}_t - \hat{\lambda}_t|] \leq \frac{\Delta}{A_t} + \sum_{i=1}^{K} \frac{\mathsf{E}_\lambda[\eta_1^i]\mathsf{E}_\lambda[m_t^i + 1]}{A_t}. \tag{4.25}$$

Then, from (4.11), (4.21) and the fact that A_t is a linear function of t we have

$$\mathsf{E}_\lambda[|\tilde{\lambda}_t - \hat{\lambda}_t|] \leq \frac{\Theta(\Delta)}{t} + \sum_{i=1}^{K} \frac{\mathsf{E}_\lambda[\eta_1^i]}{\Theta(\Delta^i)}. \tag{4.26}$$

If some Δ^i is fixed as $t \to \infty$, the second term in the right-hand side of (4.26) does not go to 0 (unless $h \to 0$, in which case $\mathsf{E}_\lambda[\eta_1^i] \to 0$ for every $1 \leq i \leq K$, due to (4.21)). However, if $\Delta^i \to \infty$ so that $\Delta^i = o(t)$ for every $1 \leq i \leq K$, then both terms in the right-hand side of (4.26) go to 0 for any given sampling period, $h > 0$, which completes the proof. □

The proof of Theorem 4.1 suggests that the proposed estimator is not consistent when both $\{\Delta^i, 1 \leq i \leq K\}$ and h are held fixed. In other words, it is necessary to have either a high sampling rate ($h \to 0$) in order to reduce the size of the unobserved overshoots or a low communication rate in all sensors ($\Delta^i \to \infty\ \forall 1 \leq i \leq K$) in order to reduce their accumulation rate.

However, an asymptotically low communication rate is not sufficient in order to preserve the asymptotic optimality of $\tilde{\lambda}_t$ in the case of discrete sampling at the sensors. For this, the sampling period h must converge to 0 at an appropriate rate relative to the communication rate and the horizon of observations, which we specify in the following theorem.

THEOREM 4.2. *If $t, \Delta^i \to \infty$ and $h \to 0$ so that*

$$\Delta^i = o(\sqrt{t}) \quad \text{and} \quad \sqrt[3]{h} = o(\Delta^i/\sqrt{t}) \qquad \forall 1 \leq i \leq K,$$

then $\sqrt{A_t}(\tilde{\lambda}_t - \lambda) \to \mathcal{N}(0, 1)$, that is, $\tilde{\lambda}_t$ is an asymptotically optimal estimator.

PROOF. Since $\sqrt{A_t}(\hat{\lambda}_t - \lambda) \sim \mathcal{N}(0, 1)$, it suffices to show that $\sqrt{A_t}|\tilde{\lambda}_t - \hat{\lambda}_t|$ converges to 0 in probability. Indeed, from (4.26) and the fact that A_t is a linear function of t,

$$\sqrt{A_t}\mathsf{E}_\lambda[|\tilde{\lambda}_t - \hat{\lambda}_t|] \leq \frac{\Theta(\Delta)}{\sqrt{t}} + \sum_{i=1}^{K} \frac{\mathsf{E}_\lambda[\eta_1^i]}{\Theta(\Delta^i/\sqrt{t})}. \tag{4.27}$$

The first term in the right-hand side goes to 0 if $\Delta = o(\sqrt{t})$. The second term goes to 0 if $\mathsf{E}_\lambda[\eta_1^i] = o(\Delta^i/\sqrt{t})$ for every $1 \le i \le K$. For the latter, it suffices that $\sqrt[3]{h} = o(\Delta^i/\sqrt{t})$ for every $1 \le i \le K$, due to (4.21), which completes the proof. $\square$

REMARK. If each Δ^i is fixed as $t \to \infty$, then Theorem 4.2 implies that $\hat{\lambda}_t$ is asymptotically efficient as $t \to \infty$ and $h \to 0$ so that $\sqrt[3]{h}\sqrt{t} \to 0$.

5. Conclusions.

In this work we considered a parameter estimation problem assuming that the statistician collects data from dispersed sensors, which observe continuous (possibly correlated) semimartingales with linear drifts with respect to a common, unknown parameter. Motivated by sensor network applications, which are typically characterized by limited communication bandwidth, we required that the sensors must send a small number of bits per transmission and that they should avoid a high rate of communication with the fusion center.

We proposed a novel methodology for this problem, according to which the sensors transmit to the fusion center one-bit messages at first exit times of appropriate statistics that they observe locally. The fusion center then combines these messages and constructs an estimator that imitates the optimal centralized estimator (which can be computed only if there is full access to the sensor observations).

We proved that the resulting estimator is consistent and, for a large class of processes, asymptotically optimal, in the sense that it attains the performance of the optimal centralized estimator when a sufficiently large horizon of observations is available. However, it is much more efficient from a practical point of view, as it reduces dramatically the congestion in the network and the computational burden at the fusion center. This is the case because it requires the transmission of only one-bit messages from the sensors and its statistical properties are preserved even with an asymptotically low rate of communication.

It remains an open problem to design estimators with analogous optimality properties in more complicated setups, such as when there is not an explicit form for the optimal centralized estimator, the dimensionality of the parameter space is large or the sensors take non-i.i.d., discrete-time observations.

Acknowledgments. The author would like to thank Dr. George V. Moustakides and Dr. Alexandra Chronopoulou for their feedback. Moreover, the author is grateful to the two anonymous referees and the Associate Editor for their valuable remarks and suggestions that led to a significant improvement of earlier versions of this work.

REFERENCES

[1] BLUM, R. S., KASSAM, S. A. and POOR, H. V. (1997). Distributed detection with multiple sensors: Part II-advanced topics. *Proc. IEEE* **85** 64–79.

[2] BROWN, B. M. and HEWITT, J. I. (1975). Asymptotic likelihood theory for diffusion processes. *J. Appl. Probab.* **12** 228–238. MR0375693

[3] BROWN, B. M. and HEWITT, J. I. (1975). Inference for the diffusion branching process. *J. Appl. Probab.* **12** 588–594. MR0378307

[4] FEIGIN, P. D. (1976). Maximum likelihood estimation for continuous-time stochastic processes. *Adv. in Appl. Probab.* **8** 712–736. MR0426342

[5] FELLOURIS, G. and MOUSTAKIDES, G. V. (2011). Decentralized sequential hypothesis testing using asynchronous communication. *IEEE Trans. Inform. Theory* **57** 534–548. MR2814070

[6] FORESTI, G. L., REGAZZONI, C. S. and VARSHNEY, P. K., eds. (2003). *Multisensor Surveillance Systems: The Fusion Perspective*. Kluwer Academic, Dordrecht.

[7] GALTCHOUK, L. and KONEV, V. (2001). On sequential estimation of parameters in semimartingale regression models with continuous time parameter. *Ann. Statist.* **29** 1508–1536. MR1873340

[8] GRENANDER, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.* **1** 195–277. MR0039202

[9] HAN, T. S. and AMARI, S. (1995). Parameter estimation with multiterminal data compression. *IEEE Trans. Inform. Theory* **41** 1802–1833.

[10] HAN, T. S. and AMARI, S. (1998). Statistical inference under multiterminal data compression. *IEEE Trans. Inform. Theory* **44** 2300–2324. MR1658791

[11] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940

[12] KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, London. MR2144185

[13] LIPTSER, R. S. and SHIRYAEV, A. N. (2001). *Statistics of Random Processes: Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **6**. Springer, Berlin. MR1800858

[14] LORDEN, G. (1970). On excess over the boundary. *Ann. Math. Statist.* **41** 520–527. MR0254981

[15] LUO, Z.-Q. (2005). Universal decentralized estimation in a bandwidth constrained sensor network. *IEEE Trans. Inform. Theory* **51** 2210–2219. MR2235295

[16] MARTINSEK, A. T. (1981). A note on the variance and higher central moments of the stopping time of an SPRT. *J. Amer. Statist. Assoc.* **76** 701–703. MR0629754

[17] MEL'NIKOV, A. V. and NOVIKOV, A. A. (1988). Sequential inferences with guaranteed accuracy for semimartingales. *Teor. Veroyatn. Primen.* **33** 480–494. MR0968395

[18] NOVIKOV, A. A. (1972). Sequential estimation of the parameters of processes of diffusion type. *Mat. Zametki* **12** 627–638. MR0317493

[19] PRAKASA RAO, B. L. S. (1985). *Statistical Inference for Diffusion Type Processes*. Arnold, London.

[20] RABI, M., MOUSTAKIDES, G. V. and BARAS, J. S. (2012). Adaptive sampling for linear state estimation. *SIAM J. Control Optim.* **50** 672–702. MR2914225

[21] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357

[22] STRIEBEL, C. T. (1959). Densities for stochastic processes. *Ann. Math. Statist.* **30** 559–567. MR0104330

[23] VEERAVALLI, V. V. (1999). Sequential decision fusion: Theory and applications. *J. Franklin Inst.* **336** 301–322. MR1674584

[24] VISWANATHAN, R. and VARSHNEY, R. K. (1997). Distributed detection with multiple sensors: Part II-fundamentals. *Proc. IEEE* **85** 54–63.

[25] XIAO, J.-J. and LUO, Z.-Q. (2005). Decentralized estimation in an inhomogeneous sensing environment. *IEEE Trans*. *Inform*. *Theory* **51** 3564–3575.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
3620 SOUTH VERMONT AVE.
KAP 416
LOS ANGELES, CALIFORNIA 90089-2532
USA
E-MAIL: fellouri@usc.edu