Automatica 72 (2016) 100–107

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Finite-horizon Gaussianity-preserving event-based sensor scheduling in Kalman filter applications[☆]

Junfeng Wu[a,b], Xiaoqiang Ren[a], Duo Han[a,1], Dawei Shi[c], Ling Shi[a]

[a] Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong

[b] ACCESS Linnaeus Center, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden

[c] State Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, PR China

ARTICLE INFO

Article history:
Received 3 April 2015
Received in revised form
1 March 2016
Accepted 2 May 2016
Available online 18 July 2016

Keywords:
Networked control systems
Estimation
Kalman filtering
Sensor scheduling
Dynamic programming

ABSTRACT

This paper considers a remote state estimation problem, where a sensor measures the state of a linear discrete-time system. The sensor has computational capability to implement a local Kalman filter. The sensor-to-estimator communications are scheduled intentionally over a finite time horizon to obtain a desirable tradeoff between the state estimation quality and the limited communication resources. Compared with the literature, we adopt a Gaussianity-preserving event-based sensor schedule bypassing the nonlinearity problem met in threshold event-based polices. We derive the closed-form of minimum mean-square error (MMSE) estimator and show that, if communication is triggered, the estimator cannot do better than the local Kalman filter, otherwise, the associated error covariance, is simply a sum of the estimation error of the local Kalman filter and the performance loss due to the absence of communication. We further design the scheduler's parameters by solving a dynamic programming (DP) problem. The computational overhead of the DP problem is less sensitive to the system dimension compared with that of existing algorithms in the literature.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The concept of controlled communication for the state estimation of a dynamical system has been prevailing in recent years. Controlled communication in general refers to reducing the communication rate intentionally to obtain a desirable tradeoff between the state estimation quality and the limited communication resources. This is rooted in the fact that the communication between the wireless sensors and the estimator at full rate is unlikely to occur for most practical applications. For instance, since the sensors are usually battery-powered and sparsely deployed, the replacement of onboard battery is not possible in most occasions. Reducing the communication rate is reasonably an alternative approach to resolve the energy saving problem. Another incentive for controlled communication is to avoid traffic congestion of the network shared by a vast number of sensors.

Estimation error covariance is most widely used for measuring the estimation quality. To minimize inevitable enlarged estimation error covariance due to the reduced communication rate, a communication scheduling strategy for a sensor is needed. Yang and Shi (2011) provided an insight that communications should be initiated periodically or more generally, as uniformly as possible, to minimize the average error covariance. For the so-called variance-based triggered scheduling in Trimpe and D'Andrea (2014), covariance recursion asymptotically converges to a periodic one. Informally, purely using the information in the error covariance is likely to lead to a periodic communication schedule. Another line of research direction such as Han, Cheng, Chen, and Shi (2013), Shi, Chen, and Darouach (2016), Shi, Chen, and Shi (2015) and Shi, Elliott, and Chen (2016) is the event-based sensor scheduling, where communication is triggered by a certain event defined on the system state. Threshold event-based communication schedules have been proposed by Battistelli, Benavoli, and Chisci (2012), Lipsa and Martins (2011), Molin (2014), Wu, Jia, Johansson, and Shi (2013) and Xu and Hespanha (2005), in different contexts but can hardly generate closed-form of the minimum mean-square error (MMSE) estimates. To obtain a tractable and simple

☆ The work by X. Ren, D. Han and L. Shi was supported by a HK RGC theme-based project T23-701/14N. The work by D. Shi was supported by the National Natural Science Foundation of China under Grant 61503027. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Gianluigi Pillonetto under the direction of Editor Torsten Söderström.

E-mail addresses: junfengw@kth.se (J. Wu), xren@connect.ust.hk (X. Ren), dhanaa@connect.ust.hk (D. Han), dshi@ualberta.ca (D. Shi), eesling@ust.hk (L. Shi).

1 Tel.: +65 8629 2706; fax: +852 2358 1485.

http://dx.doi.org/10.1016/j.automatica.2016.05.013
0005-1098/© 2016 Elsevier Ltd. All rights reserved.

estimator, Han et al. (2015) proposed a stochastic event-based mechanism, bypassing the nonlinear problem met in threshold event-based policies.

In this work we focus on a finite-horizon sensor communication scheduling problem. The sensor, as a smart one, has computational capability to implement a local Kalman filter. The utilization of the onboard computation unit has been shown to help improve estimation performance (Hovareshti, Gupta, & Baras, 2007). To alleviate the degradation of estimation performance, we adopt an event-based sensor scheduling mechanism. The benefit we obtain from this type of mechanisms is attributed to the fact that the absence of triggering provides side information to the estimator. If we pursue an optimal event-based law, it is very likely that the Gaussianity of the conditional distribution in the system state will be destroyed, for which no closed-form expression of the MMSE estimate can be derived. The distribution propagation turns out to be computationally costly under non-Gaussian circumstances. In summary, the information contained in the absence of triggering, on one hand, mitigates the Kalman filtering's performance degradation, but on the other hand, may cause difficulty in computing distribution propagation. To tackle the challenge, in this paper we introduce a similar stochastic event-based mechanism used in Han et al. (2015). Compared with Han et al. (2015), the contribution of this paper is summarized as follows:

(1) We use a simple static parameter estimation example to motivate the stochastic event-based scheduling policy. In the example, the stochastic strategy maintains the *a posteriori* distributions Gaussian with possibly the least variance.
(2) We present a closed-form expression of the MMSE estimate for the remote estimator and show that, if and when the communication is triggered, the estimator cannot do any better than the local Kalman filter, otherwise, the associated error covariance, is simply a sum of the estimation error of the local Kalman filter and the performance loss due to the absence of communication.
(3) The sensor scheduling problem can be modeled as a decision process. The sensor can sequentially design the scheduler's parameters by solving a dynamic programming (DP) problem, efficiently allocating communication resource over a finite time-horizon. The computational overhead of the DP problem is less sensitive to the dimension of systems compared with the existing works.

Notation: $\mathbb{N}$ is the set of positive integers numbers. $\mathbb{S}^n_+$ is the set of n by n symmetric positive semi-definite matrices over the real field. The notation $p(\mathbf{x}, x)$ represents the probability density function (pdf) of a random variable $\mathbf{x}$ taking value at x. For a matrix X, we abuse the notations $\det(X)$ and X^{-1}, in case of a singular matrix X, to respectively denote the pseudo-determinant and the Moore–Penrose pseudoinverse of X. The notation $X^{1/2}$ is the square root of a positive semidefinite matrix X. For a Borel set $\mathcal{B}$, $\mathfrak{L}(\mathcal{B})$ stands for the Lebesgue measure. $\times$ denotes Cartesian product and $\oplus$ stands for Minkowski addition of two sets, respectively. Define the function h: $\mathbb{S}^n_+ \mapsto \mathbb{S}^n_+$ as $h(X) \triangleq AXA' + Q$.

2. Kalman filter under controlled communication

Consider a linear time-invariant system:

$$x_{k+1} = Ax_k + w_k, \tag{1a}$$
$$y_k = Cx_k + v_k, \tag{1b}$$

where $x_k \in \mathbb{R}^n$ is the system state vector and $y_k \in \mathbb{R}^m$ is the observation vector. The noises $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are zero-mean Gaussian random vectors with $\mathbb{E}[w_k w_j'] = \delta_{kj} Q$ ($Q \geq 0$), $\mathbb{E}[v_k v_j'] = \delta_{kj} R$ ($R > 0$), where δ_{kj} is the Kronecker delta function with $\delta_{kj} = 1$ if $k = j$ and 0 otherwise, and $\mathbb{E}[w_k v_j'] = 0\ \forall j, k$. The initial state x_0 is a zero-mean Gaussian random vector that is uncorrelated with w_k and v_k and has covariance $\Sigma_0 \geq 0$. The pair (C, A) is assumed to be observable and $(A, Q^{1/2})$ is controllable.

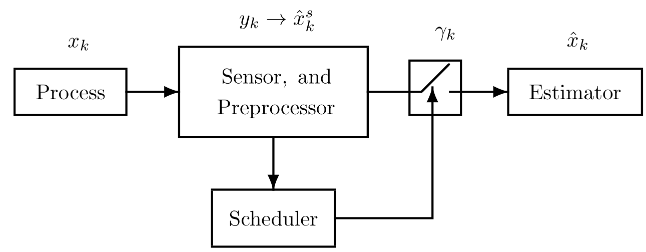


Fig. 1. Remote state estimation with a communication scheduler.

All the measurements collected by the sensor up to time k is denoted by $y_{1:k} \triangleq \{y_1, \ldots, y_k\}$. The sensor locally computes $\hat{x}^s_k \triangleq \mathbb{E}[x_k|y_{1:k}]$, the MMSE estimate of x_k based on $y_{1:k}$. Let P^s_k be the associated estimation error covariance matrix, i.e., $P^s_k \triangleq \mathbb{E}[(x_k - \hat{x}^s_k)(x_k - \hat{x}^s_k)'|y_{1:k}]$, which is computed via a standard Kalman filter initialized with $\hat{x}^s_0 = 0$ and $P^s_0 = \Sigma_0$. The sensor is equipped with a transmission scheduler (see Fig. 1), which determines whether or not $\hat{x}^s_k$ should be sent to the estimator, according to the history of transmission decision actions and the measurements collected by the sensor up to time k. Let $\gamma_k \in \{0, 1\}$ denotes the communication decision made by the scheduler. If $\gamma_k = 1$, $\hat{x}^s_k$ is sent; otherwise $\hat{x}^s_k$ is not sent. Since the sensor local estimation is initialized with $\hat{x}^s_k = 0$, without loss of generality, we assume $\gamma_0 = 1$. To focus on the role of the sensor scheduler in achieving a desired tradeoff between the remote estimation quality and communication resource, other aspects of imperfect communication, such as packet dropouts, delays and data quantization, will not be taken into account, that is, if sent by the senor, the data will reach the estimator side.

It should be noted that before deciding γ_k at time k, $\gamma_{1:k-1} \triangleq \{\gamma_1, \ldots, \gamma_{k-1}\}$ is known by the sensor. Besides, the sensor has all the measurements collected by itself. The information pattern of the sensor up to after communication at time k, if any, is denoted as $\mathcal{I}^S_k$, i.e.,

$$\mathcal{I}^S_k \triangleq \{y_1, \ldots, y_k\} \cup \{\gamma_1, \ldots, \gamma_k\}, \quad \text{with } \mathcal{I}^S_0 = \emptyset.$$

Similarly, we denote by $\mathcal{I}^E_k$ the information pattern at the remote estimator up to after communication at time k. Because of the perfect communication channel assumed, γ_k is known to the estimator. $\mathcal{I}^E_k$ contains both the history of communication actions $\gamma_{1:k}$ and the measurement data received from the sensor, that is,

$$\mathcal{I}^E_k = \{\gamma_1 \hat{x}^s_1, \ldots, \gamma_k \hat{x}^s_k\} \cup \{\gamma_1, \ldots, \gamma_k\}, \quad \text{with } \mathcal{I}^E_0 = \emptyset.$$

We define a communication scheduling policy applied by the sensor at time k as a function f_k:

$$\gamma_k = f_k(\mathcal{I}^S_{k-1}, y_k), \tag{2}$$

where f_k's are assumed to be measurable mappings. A finite-horizon sensor communication policy Θ is accordingly defined as a sequence of f_k's: $\Theta \triangleq \{f_1, f_2, \ldots, f_T\}$. Because the estimator is aware of Θ being used by the sensor, it computes $\tilde{x}_k$, its own estimate of the state x_k based on $\mathcal{I}^E_k$, $\tilde{x}_k = g_k(\mathcal{I}^E_k)$, where g_k's are measurable mappings. A finite-horizon remote estimator Ξ is accordingly defined as a sequence of g_k's: $\Xi \triangleq \{g_1, g_2, \ldots, g_T\}$. The estimator computes P_k, the corresponding estimation error covariance matrix, as: $P_k = \mathbb{E}_\Theta\left[(x_k - \tilde{x}_k)(x_k - \tilde{x}_k)' \mid \mathcal{I}^E_k\right]$, where $\mathbb{E}_\Theta[\cdot|\cdot]$ denotes conditional expectation with respect to a fixed Θ.

Let us denote by $J_{\mathscr{E}}(\Theta, \Xi)$ the penalty associated with the total expected estimation errors of the remote estimator within a horizon $\mathtt{T}$, i.e.,

$$J_{\mathscr{E}}(\Theta, \Xi) = \mathbb{E}\left[\sum_{k=1}^{\mathtt{T}} \mathrm{Tr}\left(\mathbb{E}_{\Theta}\left[(x_k - \tilde{x}_k)(\cdot)' | \mathcal{I}_k^{\mathrm{E}}\right]\right)\right]. \tag{3}$$

The communication cost incurred is defined as the expected sensor-to-estimator communication times within the horizon $\mathtt{T}$: $J_{\mathscr{C}}(\Theta) = \mathbb{E}\left[\sum_{k=1}^{\mathtt{T}} \mathbb{E}_{\Theta}[\gamma_k | \mathcal{I}_{k-1}^{\mathrm{E}}]\right]$. The expectation is taken with respect to the noise processes w_k's and v_k's. We simultaneously consider $J_{\mathscr{E}}(\Theta, \Xi)$ and $J_{\mathscr{C}}(\Theta)$ within a single objective function:

$$\underset{\Theta, \Xi}{\text{minimize}} \; J(\lambda, \Theta, \Xi) \triangleq J_{\mathscr{E}}(\Theta, \Xi) + \lambda J_{\mathscr{C}}(\Theta), \tag{4}$$

where λ is a Lagrange coefficient. A range of rate-error tradeoff problems, such as Imer and Başar (2005), Li, Lemmon, and Wang (2010), Molin (2014), Wu, Johansson, and Shi (2014), etc., are incorporated in (4).

In general, it is difficult to find a global optimal solution of (4) since Θ and Ξ are coupled in $J_{\mathscr{E}}(\Theta, \Xi)$. We first give a quick answer to "what is the optimal estimator Ξ^* for problem (4)?"by invoking the following lemma.

Lemma 1. *For any given Θ, the optimal estimator Ξ^* for the estimator is the MMSE estimator $\hat{x}_k \triangleq g_k^*(\mathcal{I}_k^{\mathrm{E}}) = \mathbb{E}_{\Theta}[x_k | \mathcal{I}_k^{\mathrm{E}}]$.*

Proof. When Θ is fixed, $J_{\mathscr{C}}(\Theta)$ in (4) is fixed and only $J_{\mathscr{E}}(\Theta, \Xi)$ is to be minimized. Then, from the optimal filtering theory Anderson and Moore (1979), Ξ^* is given by the MMSE estimator of x_k, which is also specified as the conditional expectation of x_k given $\mathcal{I}_k^{\mathrm{E}}$. ■

Having established Lemma 1, we only need to focus on seeking an optimal communication scheduling policy Θ^*. However, Problem (4) remains challenging, as Ξ^* is affected by the choice of Θ through the *a posteriori* pdf $p_{\Theta}(x_k, x | \mathcal{I}_k^{\mathrm{E}})$. Even if this problem has been relaxed to a Markov decision process (MDP) by various means in the literature, e.g., an iterative algorithm was presented in Molin (2014), and a fixed estimator was imposed in Gatsis, Ribeiro, and Pappas (2013) and Li et al. (2010), it may still encounter computational difficulties. When fixing a Ξ and executing an algorithm to solve the MDP, one resorts to solving a Bellman equation backwards:

$$u_k(s_k) = \min_{\gamma_k = \{0,1\}} \left\{ c_{\Xi}(s_k, \gamma_k) + \mathbb{E}_{\Xi}[u_{k+1}(s_{k+1}) | s_k, \gamma_k] \right\}, \tag{5}$$

where $s_k \in \mathcal{S}$ is the state of the MDP, $c_{\Xi}(s_k, \gamma_k)$ is the cost involved, given Ξ, when the system is in state s_k and performs action γ_k, and $u_{\mathtt{T}}(s_k) = 0$ for any $s_k \in \mathcal{S}$. In each of the aforementioned works, although differently defined, but in general, s_k is a variable related to x_k. To numerically solve (5), a first step is to discretize the continuous state space. Decisions should be made to select an action for each state in the discretized state space. In order to guarantee the desired calculation accuracy, computational complexity exponentially increases with respect to the dimension of system (1a), which makes it impossible to scale the algorithm to high-dimensional systems. In addition, the MDPs formulated in the aforementioned works lead to nonlinear state estimation problems, for which a closed-form expression of the MMSE estimate is difficult to obtain. To obtain an exact MMSE estimate, it is required to calculate the *a posteriori* pdf in the state x_k at each time, using Bayes' theorem. This recursive propagation only has a conceptual solution and cannot be computed analytically. Gaussian approximation is a suboptimal solution to solve the intractable nonlinear filtering problem heuristically. Interested readers can refer to Shi, Chen, and Shi (2014), Sijs and Lazar (2012) and Wu et al. (2013) for threshold event-triggering. In their works, numerical integrations and approximations were involved. However, after a few updates, the *a posteriori* distribution may become a poor approximation.

3. Gaussianity-preserving event-based sensor scheduling for state estimation

The beauty of the event-based communication policy for state estimation is its resources saving, but it also encounters analytical and computational difficulties when it is deterministic as we discussed in the previous section. To overcome the computational complexity issue and the nonlinear state estimation problem, we adopt a special type of communication schedules that can preserve Gaussianity of the *a posteriori* pdf in the system state. By taking the advantage of Gaussianity, the computational complexity issue and the nonlinearity of the state estimation can be bypassed.Parameters of such a scheduling policy are to be designed to fulfill performance restrictions.

The event-based communication scheduling to be adopted was originated in Han et al. (2015). While a primitive sensor is considered in Han et al. (2015), a modified policy for a smart sensor case is presented in this work. To introduce this policy, let us first define z_k, the incremental innovative information of $\hat{x}_k^s$ relative to the information sent by the sensor at the most recent instance before time k, as

$$z_k \triangleq \hat{x}_k^s - A^{\tau_k - 1} \hat{x}_{\mathtt{N}_{k-1}}^s \tag{6}$$

where $\mathtt{N}_k \triangleq \max\left\{ j : \gamma_j = 1, 1 \le j \le k \right\}$ denotes the most recent triggering instance before time k and $\tau_k \triangleq k - \mathtt{N}_k + 1$ represents the distance between $k+1$ and $\mathtt{N}_k$. Both $\mathtt{N}_k$ and τ_k are measurable to $\mathcal{I}_k^{\mathrm{E}}$. We now present the Gaussianity-preserving communication scheduling policy $\widehat{\Theta}$ as follows. At each time k, the sensor randomly generates a uniformly distributed random variable $\xi_k \in [0, 1]$. Let

$$\gamma_k = \begin{cases} 0, & \text{if } \xi_k \le s(z_k, \Gamma_k), \\ 1, & \text{otherwise}, \end{cases} \tag{7}$$

where $s(\cdot, \cdot)$ is defined as

$$s(x, \Gamma) = e^{-\frac{1}{2} x' \Gamma^{-1} x}, \quad x \in \mathbb{R}^n \text{ and } \Gamma \in \mathbb{S}_+^n, \tag{8}$$

and $\Gamma_k \in \mathbb{S}_+^n$ is a weight matrix to be designed. At each time communication is randomly triggered with probability $1 - s(z_k, \Gamma_k)$. If $s(z_k, \Gamma_k)$ is smaller, the sensor-to-estimator communication is more likely to be triggered.

Remark 1. The realization γ_k not only depends on the outcome of z_k but also on an auxiliary random variable ξ_k. Though introducing a little more uncertainty, ξ_k plays a crucial role of smoothing the *a posteriori* distribution into a Gaussian one when $\gamma_k = 0$ as we will see later. Some other type of strategies might produce the same effect, but to the best of our knowledge, there are no any such deterministic ones in the literature.

The scheduling policy (7) does not produce as good estimation performance as deterministic ones, since ξ_k introduces additional uncertainty. However, the following simple example suggests that the performance loss is possibly the least, where a static parameter estimation problem is given to help readers understand the idea in an easier way. After this example, in the rest of this section, we will provide answers to the following two questions:

(1) How shall we compute the remote estimate and the associated estimation error covariance?
(2) How shall we design the parameters for $\widehat{\Theta}$ to minimize $J(\lambda, \widehat{\Theta}, \Xi)$?

3.1. State parameter estimation example

Consider a static parameter estimation problem: a parameter $\mathbf{x}$ is observed by a set of sensors indexed by $\mathcal{V} = \{1, \ldots, l\}$ via

$$\mathbf{y}^{(i)} = \mathbf{x} + \mathbf{v}^{(i)}, \quad i \in \mathcal{V}, \tag{9}$$

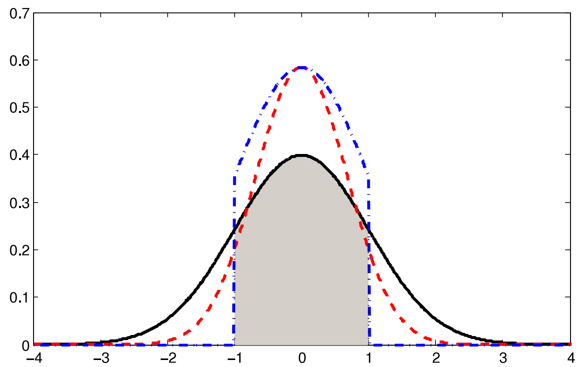


Fig. 2. Illustrations of a Gaussian pdf (black full line), a truncated Gaussian pdf (red dash-dotted line), and a Gaussian pdf majorized by the truncated Gaussian pdf (red dash line).

where $\mathbf{y}^{(i)}, \mathbf{x}, \mathbf{v}^{(i)} \in \mathbb{R}$, and $p(\mathbf{x}, x) \sim \mathcal{N}(\mu, \sigma_{\mathbf{x}})$ and $p(\mathbf{v}^{(i)}, v) \sim \mathcal{N}(0, \sigma_{\mathbf{v}}^{(i)})$ are independent of each other. Then $p(\mathbf{y}^{(i)}, y) \sim \mathcal{N}(\mu, \sigma_{\mathbf{y}}^{(i)})$ with $\sigma_{\mathbf{y}}^{(i)} = \sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(i)}$. Without loss of generality, we let $\mu = 0$. All the measurements are scheduled to be sent to an estimator. Let $\gamma^{(i)}$ denote whether or not $\mathbf{y}^{(i)}$ is sent: $\gamma^{(i)} = 1$ if $\mathbf{y}^{(i)}$ is sent; otherwise $\gamma^{(i)} = 0$. A commonly used scheduling policy, denoted as θ, is a threshold one, which is of the following form:

$$\gamma^{(i)} = \begin{cases} 0, & \text{if } \dfrac{|\mathbf{y}^{(i)} - \hat{\mathbf{x}}^{(i-1)}(\theta)|}{\sigma_{\mathbf{y}}^{(i)}} \le \delta_i, \\ 1, & \text{otherwise,} \end{cases} \tag{10}$$

where $\hat{\mathbf{x}}^{(i-1)}(\theta)$ is the MMSE estimate of $\mathbf{x}$ based on $\{\gamma^{(1)}, \dots, \gamma^{(i-1)}\} \cup \{\gamma^{(1)}\mathbf{y}^{(1)}, \dots, \gamma^{(i-1)}\mathbf{y}^{(i-1)}\}$ with $\hat{\mathbf{x}}^{(0)}(\theta) = 0$ and δ_i is a parameter to be designed to fulfill the tradeoff between communication cost and estimation error. Measurements from the l number of sensors, if sent, are collected in numerical order and fused.

Suppose $\gamma^{(1)} = 0$. The conditional pdf of $\mathbf{y}^{(1)}$ is

$$p_\theta(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0) = \begin{cases} \dfrac{p(\mathbf{y}^{(1)}, y)}{1 - 2q(\delta_0)}, & \text{if } \gamma^{(1)} = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where $q(x) \triangleq \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_0^2}{2}} \mathrm{d}x_0$ is the Q-function. It is evident that $p_\theta(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0)$ is not Gaussian. Fig. 2 illustrates $p_\theta(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0)$ when $\sigma_{\mathbf{y}} = 1$. Conditioning on $\mathbf{y}^{(1)}$, the MMSE estimate of $\mathbf{x}$ is $\hat{\mathbf{x}}^{(1)}(\theta) \triangleq \frac{\sigma_{\mathbf{x}}}{\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)}} \mathbf{y}^{(1)}$. From optimal filtering theory, $\mathbf{x} - \hat{\mathbf{x}}^{(1)}(\theta)$ is orthogonal to $\mathbf{y}^{(1)}$. Noting that $\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}^{(1)}(\theta) + \frac{\sigma_{\mathbf{x}}}{\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)}} \mathbf{y}^{(1)}$ and $\mathbf{x} - \hat{\mathbf{x}}^{(1)}(\theta) \sim \mathcal{N}(0, \sigma_{\mathbf{x}} - \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)}})$, through $p_\theta(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0)$, we obtain that $p_\theta(\mathbf{x}, x|\gamma^{(1)} = 0)$ *is not Gaussian.* Then, when $\mathbf{y}^{(2)}$ (or $\gamma^{(2)} = 0$) is received, the conditional distribution is updated according to Bayes' theorem. Due to loss of Gaussianity, updating the conditional distribution at each step needs to resort to numerical integrations, which is computationally intractable. To overcome the computational problem, let $\gamma^{(1)}$ be decided according to an analogue, denoted by $\hat{\theta}$, of policy (7):

$$\gamma^{(1)} = \begin{cases} 0, & \text{if } \xi^{(1)} \le s(\mathbf{y}^{(1)}, \Gamma^{(1)}), \\ 1, & \text{otherwise,} \end{cases} \tag{12}$$

where $\xi^{(1)}$ is a uniformly distributed random variable in $[0, 1]$, $s(\cdot, \cdot)$ is defined as (8) and $\Gamma^{(1)} \in \mathbb{R}$ is a weight.

Lemma 2. *Consider system (9) and scheduling policy $\hat{\theta}$ given in (12) for $\gamma^{(1)}$. Then $p_\theta(\mathbf{x}, x|\gamma^{(1)} = 0)$ is Gaussian. Moreover, suppose δ_1 and $\Gamma^{(1)}$ are appropriately designed such that $\mathbb{P}_\theta(\gamma^{(1)} = 0) = \mathbb{P}_{\hat{\theta}}(\gamma^{(1)} = 0)$, then $p_{\hat{\theta}}(\mathbf{x}, x|\gamma^{(1)} = 0) \sim \mathcal{N}(0, \sigma_{\mathbf{x}} - \frac{4q(\delta_1)(1-q(\delta_1))\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)}})$.*

The proof of Lemma 2, which is presented in the Appendix, is reported after the proof of Lemma 4, since some techniques are borrowed from the latter one. If we use scheduling policy analogous to (12) for each fusion, the *a posteriori* pdfs remain Gaussian, reducing the estimation problem from tracking of distributions, which is usually computationally intractable, to tracking of the mean and the covariance of distributions. Since the scheduling policy $\hat{\theta}$ is randomized, it can be expected that $p_{\hat{\theta}}(\mathbf{x}, x|\gamma^{(1)} = 0)$ is more disordered and has a larger covariance compared with $p_\theta(\mathbf{x}, x|\gamma^{(1)} = 0)$. Here the "disorder" can be interpreted by the idea of majorization (cf., Lipsa & Martins, 2011).

Definition 1. Suppose p and p_* are two pdfs on $\mathbb{R}$. We say p majorizes p_*, which is denoted by $p \succ p_*$, if for any Borel set $\mathcal{B}_1 \in \mathbb{R}$ with $\mathfrak{L}(\mathcal{B}_1) < \infty$, there always exists another Borel set $\mathcal{B}_2 \in \mathbb{R}$ satisfying $\mathfrak{L}(\mathcal{B}_1) = \mathfrak{L}(\mathcal{B}_2)$ and $\int_{\mathcal{B}_2} p \,\mathrm{d}\mu \ge \int_{\mathcal{B}_1} p_* \,\mathrm{d}\mu$.

The following statement suggests that, although leading to some performance loss, $\hat{\theta}$ generates the *a posteriori* pdf with possibly the least variance.

Lemma 3. *Consider system (9) and scheduling strategies $\hat{\theta}$ and θ for $\gamma^{(1)}$. Suppose δ_1 and $\Gamma^{(1)}$ are appropriately designed such that $\mathbb{P}_\theta(\gamma^{(1)} = 0) = \mathbb{P}_{\hat{\theta}}(\gamma^{(1)} = 0)$. Then $p_\theta(\mathbf{x}, y|\gamma^{(1)} = 0) \succ p_{\hat{\theta}}(\mathbf{x}, y|\gamma^{(1)} = 0)$. Moreover, if there exists a $\sigma > 0$ such that $p_\theta(\mathbf{x}, y|\gamma^{(1)} = 0) \succ \mathcal{N}(0, \sigma)$, it implies $\sigma \ge \sigma_{\mathbf{x}} - 4q(\delta_1)(1 - q(\delta_1))\sigma_{\mathbf{x}}^2/(\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)})$.*

Proof. The majorization relation between $p_\theta(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0)$ and $p_{\hat{\theta}}(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0)$ follows from the observation that $p_\theta(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0) \ge p_{\hat{\theta}}(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0)$ for any $-\sigma_{\mathbf{y}}^{(1)}\delta_1 \le y \le \sigma_{\mathbf{y}}^{(1)}\delta_1$. Since $p_\theta(\mathbf{y}^{(1)}, 0|\gamma^{(1)} = 0) = p_{\hat{\theta}}(\mathbf{y}^{(1)}, 0|\gamma^{(1)} = 0)$, it is impossible for the former pdf to majorize a Gaussian distribution having a variance smaller than that of the latter one. From optimal filtering theory, $\hat{\mathbf{x}}^{(1)}(\theta) = \frac{\sigma_{\mathbf{x}}}{\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)}} \mathbf{y}^{(1)}$ and $\mathbf{x} - \hat{\mathbf{x}}^{(1)}(\theta)$ is orthogonal to $\mathbf{y}^{(1)}$, then by Lipsa and Martins (2011, Lemma 3) the majorization relation holds under the convolution of the above pdfs, which completes the proof. ∎

The scheduling policy (12) can be extended into $\mathbb{R}^n$. However, for multi-dimensional cases, the weight matrix $\Gamma^{(i)}$ corresponding to a given communication rate is not unique.

3.2. MMSE estimation

In the following lemma, we show that, using $\widehat{\Theta}$, when the triggering is absent, the conditional distribution of x_k is still Gaussian. This property is preliminary for computing the MMSE estimate and the associated estimation error covariance.

Lemma 4. *Consider $\widehat{\Theta}$ given in (7). The conditional distribution of x_k given $\mathcal{I}_k^{\mathrm{E}}$ is Gaussian, i.e.,*

$$p_{\widehat{\Theta}}(x_k, x|\mathcal{I}_k^{\mathrm{E}}) \sim \begin{cases} \mathcal{N}(\hat{x}_k^s, P_k^s), & \text{if } \gamma_k = 1, \\ \mathcal{N}(A^{\tau_k - 1}\hat{x}_{\mathrm{N}_k}^s, P_k^s + \Psi_k), & \text{if } \gamma_k = 0, \end{cases}$$

where Ψ_k is governed by the following recursive equations:

$$\Sigma_k = (1 - \gamma_{k-1})A\Psi_{k-1}A' + h(P_{k-1}^s) - P_k^s, \tag{13}$$

$$\Psi_k = (\Sigma_k^{-1} + \Gamma_k^{-1})^{-1}, \tag{14}$$

with initial value $\Psi_0 = 0$.

It can be seen from Lemma 4 that Γ_k^{-1}, like a fisher information matrix, represents the side information gained from the absence of the communication.

Lemma 5. *Consider $\widehat{\Theta}$ given in (7). The probability of no transmission is*

$$\mathbb{P}(\gamma_k = 0|\mathcal{I}^{\mathrm{E}}_{k-1}) = \det(\Psi_k \Sigma_k^{-1})^{1/2}. \tag{15}$$

The proofs of Lemmas 4 and 5 are put in the Appendix.

The following two theorems show how the remote estimator calculates its own estimate and the corresponding estimation error covariance recursively under $\widehat{\Theta}$. The simple and efficient recursion is from the Gaussianity of the *a posteriori* distribution. The proofs are straightforward from Lemma 4.

Recall that $\hat{x}^s_k$ and P^s_k are updated locally at the sensor using a standard Kalman filter.

Theorem 1. *Consider $\widehat{\Theta}$ given in (7). The MMSE estimator Ξ^* is given by*

$$\hat{x}_k = \begin{cases} \hat{x}^s_k, & \text{if } \gamma_k = 1, \\ A\hat{x}_{k-1}, & \text{if } \gamma_k = 0, \end{cases} \tag{16}$$

with $\hat{x}_k = 0$.

For a concise notation, we define the following positive semidefinite matrix:

$$\Phi_k \triangleq (1-\gamma_k)\Psi_k, \tag{17}$$

Then Φ_k can be computed according to (13), (14) and (17).

Theorem 2. *Consider $\widehat{\Theta}$ given in (7) and the MMSE estimator Ξ^* given in Theorem 1. The estimation error covariance P_k of the estimator is computed as*

$$P_k = P^s_k + \Phi_k. \tag{18}$$

Theorems 1 and 2 show that, when the sensor is capable of implementing a local Kalman filter, the remote estimator can update its estimate and the associated estimation error covariance in a simple and efficient way. When $\gamma_k = 1$, $(\hat{x}_k, P_k)$ is updated as $(\hat{x}^s_k, P^s_k)$; while when $\gamma_k = 0$, Ψ_k is the performance loss of the remote estimator caused by the lack of a point-valued measurement. Compared with an open-loop prediction, Γ_k^{-1} in Ψ_k can be interpreted as side information gained from the absent of communication. Theorem 2 exhibits that P_k is simply a sum of P^s_k, which reflects estimation error of the sensor's local Kalman filter, and Ψ_k, which reflects the performance loss due to absence of communication.

3.3. Parameter optimization via dynamic programming

The estimation error covariance is a crucial parameter to evaluate estimation performance. To calculate it, MDP-based algorithms, such as Imer and Başar (2005), Li et al. (2010), Molin (2014) and Wu et al. (2014), need to know the *a posteriori* pdf of x_k. This results in considerable computational overhead when the dimension of system (1a) increases. In contrast, by taking the advantages of the closure of Gaussian distribution under convolutions, calculating the estimation error covariance under $\widehat{\Theta}$ is tractable. In this section, we first demonstrate that the estimation penalty $J_{\mathscr{E}}(\widehat{\Theta}, \Xi^*)$ and the communication penalty $J_{\mathscr{C}}(\widehat{\Theta})$ can be computed simply by breaking things down according to either $\gamma_k = 0$ or 1. Then we design the parameters in $\widehat{\Theta}$ by solving a DP problem.

3.3.1. Cost function

In view of Theorem 2, Φ_k is a function of $\gamma_{1:k}$ under a given $\widehat{\Theta}$ and the optimal estimator Θ^*. The estimation penalty defined in (3) can be decomposed as: $J_{\mathscr{E}}(\widehat{\Theta}) \triangleq \mathbb{E}_{\widehat{\Theta}}\left[\sum_{k=0}^{\mathrm{T}} \mathrm{Tr}\left(P^s_k + \Phi_k\right)\right]$. Since $\sum_{k=0}^{\mathrm{T}} P^s_k$ does not depend on $\widehat{\Theta}$, we ignore this term and redefine the cost function in (4) under Θ^* as follows:

$$J(\lambda, \widehat{\Theta}) \triangleq \mathbb{E}_{\widehat{\Theta}}\left[\sum_{k=1}^{\mathrm{T}} \mathrm{Tr}(\Phi_k)\right] + \lambda \mathbb{E}_{\widehat{\Theta}}\left[\sum_{k=1}^{\mathrm{T}} \gamma_k\right]. \tag{19}$$

The optimal Gaussianity-preserving communication scheduling policy $\widehat{\Theta}^*$ is defined as

$$\widehat{\Theta}^* = \arg\min_{\widehat{\Theta}} J(\lambda, \widehat{\Theta}). \tag{20}$$

3.3.2. Design of Gaussianity-preserving scheduling policy

To solve $\widehat{\Theta}^*$, we need to search Γ_k (or equivalently Ψ_k) at each time. When the dimension of system (1a) exceeds 1, Ψ_k is chosen over a positive semidefinite convex cone as long as $\Sigma_k \geq \Psi_k$ and $\mathrm{rank}(\Sigma_k) = \mathrm{rank}(\Psi_k)$. To reduce the computational complexity, we restrict the searching space within $\Psi_k = \rho_k \Sigma_k$, where $\rho_k \in (0, 1)$ is a ratio to be designed. By doing so, $\Gamma_k = (1/\rho_k - 1)^{-1}\Sigma_k$. Then, instead of dealing with $\frac{n(n+1)}{2}$ degrees of freedom searching Ψ_k, now we only need to design a scaling ratio ρ_k. Further analysis when this restriction is removed is left to future work.

When $\Psi_k = \rho_k \Sigma_k$ in $\widehat{\Theta}$, (18) is reduced to $\mathbb{P}(\gamma_k = 0|\mathcal{I}^{\mathrm{E}}_{k-1}) = (\rho_k)^{r_k/2}$, where $r_k \triangleq \mathrm{rank}(\Sigma_k)$. Then (19) can be rewritten as

$$J(\lambda, \widehat{\Theta}) = \mathbb{E}_{\widehat{\Theta}}\left[\sum_{k=0}^{\mathrm{T}-1}\left(\mathrm{Tr}(\Phi_k) + \lambda(1 - \rho_{k+1}^{r_{k+1}/2})\right) + \mathrm{Tr}(\Phi_{\mathrm{T}})\right]$$

with $\Phi_0 = 0$. The following lemma provides a method for calculating r_k offline.

Lemma 6. *For given $\mathcal{I}^{\mathrm{E}}_{k-1}$, we have*

$$r_k = \mathrm{rank}\left(h^{\tau_{k-1}}(P^s_{\mathrm{N}_{k-1}}) - P^s_k\right).$$

Proof. First observe that $r_k = \mathrm{rank}(\Sigma_k)$. By (13) and (14), Σ_k can be written as

$$\Sigma_k = \sum_{i=1}^{\tau_{k-1}} \alpha_i \left(h^i(P^s_{k-i}) - h^{i-1}(P^s_{k-i+1})\right)$$

for some certain constants $\alpha_i \in (0, 1)$. Then,

$$\begin{aligned} \mathrm{Im}(\Sigma_k) &= \mathrm{Im}\left(\sum_{i=1}^{\tau_{k-1}}\left(h^i(P^s_{k-i}) - h^{i-1}(P^s_{k-i+1})\right)\right) \\ &= \mathrm{Im}\left(h^{\tau_{k-1}}(P^s_{\mathrm{N}_{k-1}}) - P^s_k\right), \end{aligned}$$

which completes the proof. ■

We define by $s_k = \{\Phi_k, \tau_k\}$ and ρ_{k+1} the state and the action at time slot k, respectively. Recall that τ_k is measurable to $\mathcal{I}^{\mathrm{E}}_k$. Given the current state s_k and the action ρ_{k+1}, the next-step state s_{k+1} can be obtained as

$$s_{k+1} = \begin{cases} \{0, 1\}, & \text{if } \gamma_{k+1} = 1, \\ \{\rho_{k+1}\Sigma_{k+1}, \tau_k + 1\}, & \text{if } \gamma_{k+1} = 0, \end{cases} \tag{21}$$

where $\Sigma_{k+1} = A\Phi_k A' + h(P^s_k) - P^s_{k+1}$. The transition probability is determined by ρ_{k+1}. The one-stage cost function is given by

$$c(s_k, \rho_{k+1}) = \mathrm{Tr}(\Phi_k) + \lambda(1 - \rho_{k+1}^{r_{k+1}/2}).$$

Since r_{k+1} is a function of τ_k by Lemma 6, $c(s_k, \rho_{k+1})$ is fully determined by s_k and ρ_{k+1}. Therefore, the optimal ρ_k's can be found by solving a DP problem. Obviously, the action space is $\mathcal{A} =$

$(0, 1)$. The state space for Φ_k, denoted as $\hat{\mathcal{S}}_k$, can be computed in a recursive way:

$$\hat{\mathcal{S}}'_{k+1} \triangleq A \times \hat{\mathcal{S}}_k \times A' \oplus \left[h(P_k^s) - P_{k+1}^s\right],$$
$$\hat{\mathcal{S}}_{k+1} = (0, 1) \times \hat{\mathcal{S}}'_{k+1},$$

with initial value $\hat{\mathcal{S}}_0 = 0$. The state space for τ_k, denoted as $\breve{\mathcal{S}}_k$, can be obtained in a similar way as $\breve{\mathcal{S}}_{k+1} = \{\breve{\mathcal{S}}_k \oplus \{1\}, 1\}$, with initial value $\breve{\mathcal{S}}_0 = 1$. Since $\tau_k = 1$ must hold as long as $\Phi_k = 0$, the state space of the DP problem for s_k, $\mathcal{S}_k$, can be defined as

$$\mathcal{S}_k = \left\{\{\hat{\mathcal{S}}_k \setminus 0\} \times \{\breve{\mathcal{S}}_k \setminus 1\}, \{0, 1\}\right\}. \tag{22}$$

The optimal scheduling policy hence can be obtained by solving the following Bellman's equation:

$$\hat{J}_{\mathrm{T}}(s_{\mathrm{T}}) = \mathrm{Tr}(\Phi_{\mathrm{T}}),$$
$$\hat{J}_k(s_k) = \min_{\rho_{k+1}\in[0,1]} \Big\{\lambda(1-\rho_{k+1}^{r_{k+1}/2}) + \mathrm{Tr}(\Phi_k) + \rho_{k+1}^{r_{k+1}/2}\hat{J}_{k+1}\left(s_{k+1} = \{\rho_{k+1}\Sigma_{k+1}, \tau_k + 1\}\right) + (1-\rho_{k+1}^{r_{k+1}/2})\hat{J}_{k+1}(s_{k+1} = \{0, 1\})\Big\}, \tag{23}$$

where $\hat{J}_k(s_k)$ is the optimal cost-to-go function at time slot k. Note that since the action space is compact and the one stage cost function is bounded, the optimal policy exists (Puterman, 2009). The above Bellman equation (23) can be solved using standard DP techniques, such as *backward induction algorithm*.

From Cheung, Kwok, and Lau (2013), the computational complexity of computing r_k (involves computing the rank of a matrix and matrix multiplication) is less than $O\left(n^3\right)$, where n is the dimension of system (1a). It is well known that for an MDP with finite state space (of cardinality $\mathtt{N}$), finite action space (of cardinality $\mathtt{M}$) and finite horizon $\mathtt{T}$, the computational complexity is $O\left(\mathtt{N}^2\mathtt{MT}\right)$ from Tsitsiklis (2007). If the state space is continuous, it needs to be discretized first. In general, to provide a solution with desired accuracy, the computational complexity increases exponentially with the dimension of the state space (Chow & Tsitsiklis, 1989). In our algorithm, discretization of the state space is equivalent to the discretization of ρ_k, which is a scalar over $(0, 1)$ (independent of the dimension of the system (1a)). On the contrary, for the MDPs formulated in Gatsis et al. (2013), Li et al. (2010) and Molin (2014), the computation complexity scales at least exponentially with system (1a). Note that the dimension of the state space for our algorithm is independent of that of system (1a), however it scales rapidly (exponentially) as $\mathtt{T}$ increases. To alleviate the computation burden, we may set a time-out condition for communication. To be precise, the measurement will be sent if τ_k is greater than a certain threshold regardless of its content. Since if $\gamma_k = 1$, $\hat{\mathcal{S}}_k = 0$, the time-out mechanism helps trim the size of the state space.

4. Numerical examples

To illustrate the Gaussianity-preserving property of the sensor scheduling $\widehat{\Theta}$ defined in (7), we simulate the static parameter estimation circumstance (9) since the Gaussianity-preserving event-based scheduling works for parameter estimation and state estimation in the same way. Consider $\mathbf{y}^{(1)} = \mathbf{x} + \mathbf{v}^{(1)}$, with $p(\mathbf{x}, x) \sim \mathcal{N}(0, 1)$ and $p(\mathbf{v}^{(1)}, v) \sim \mathcal{N}(0, 1)$. let $\Gamma^{(1)} = 1$ in (12). A Monte Carlo experiment of 10^6 simulations is performed, and data are recorded. We partition the real line into many small intervals $(\underline{b}_i, \overline{b}_i]$ and use the following approximation: for any $y \in (\underline{b}_i, \overline{b}_i]$,

$$p_{\hat{\theta}}(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0) \approx \frac{\sharp \text{ of trails with } y \text{ in} (\underline{b}_i, \overline{b}_i] \text{ and with } \gamma^{(1)} = 0}{(\sharp \text{ of trails with } \gamma^{(1)} = 0) \times (\text{width of the interval})}.$$

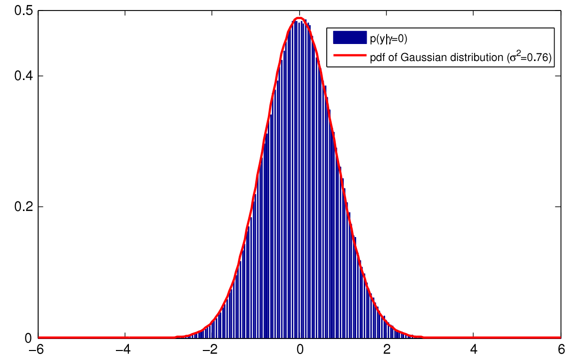


Fig. 3. Comparison between $p_{\hat{\theta}}(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0)$ and $\mathcal{N}(0, 0.76)$.

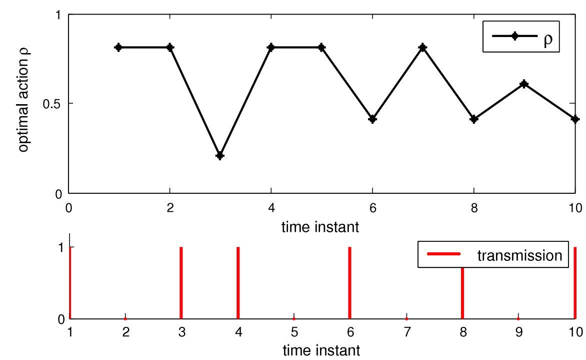


Fig. 4. Realization of the decision process in Section 3.3. The red arrows represent communications of the sensor's local information. The black dots represent the values that ρ's are chosen at each time k.

Fig. 3 shows that the numerical result matches the pdf of the Gaussian distribution $\mathcal{N}(0, 0.67)$ very well.

Then, we consider the application of Kalman filtering for system (1) with $A = \begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix}$, $C = \begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix}$, $Q = \mathrm{diag}\left([9\ 9\ 9]\right)$, $R = 9$. Furthermore, we take $P_0^s = \mathrm{diag}\left([6\ 1\ 9]\right)$.

To illustrate the communication scheduling policy, we choose $T = 10$, $\lambda = 150$, and let ρ_k's take values from the set $\{0.01, 0.2, 0.4, 0.6, 0.8, 0.99\}$. In Fig. 4, we show a realization of the decision process and the timing of communication. We also illustrate the tradeoffs between the transmission usage and the estimation performance with respect to different λ's. In Fig. 5, When λ increases, $\mathbb{E}_{\widehat{\Theta}}\left[\sum_{k=1}^T \gamma_k\right]$ reduces while $\mathbb{E}_{\widehat{\Theta}}\left[\sum_{k=1}^T \mathrm{Tr}(\Phi_k)\right]$ grows, which suggests that we should choose a relatively large λ in the presence of strict communication constraints and a relatively small one the other way round.

5. Conclusion

We considered a remote state estimation problem. In this problem, a sensor measures the state of a linear discrete-time system and runs a local Kalman filter. The sensor's local information is transmitted to the remote estimator via a communication channel. The communications are scheduled via a Gaussianity-preserving event-based sensor scheduler over a finite time horizon to save communication resources. By doing so, the estimator can gain side information from absence of communication, and moreover, the *a posteriori* distributions in system states are smoothed into Gaussian ones. We derived the optimal estimator and communication rate under such a

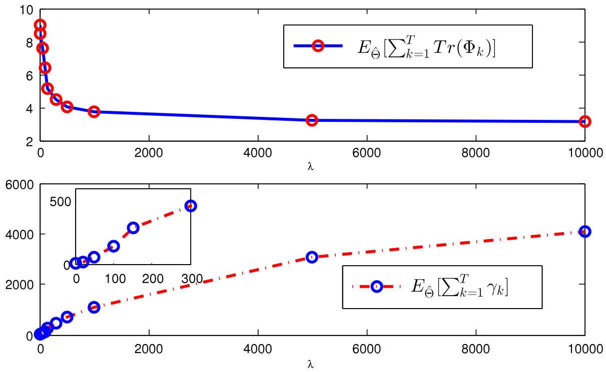


Fig. 5. Tradeoffs between the communication usage and the estimation quality. When λ increases, the total communication times reduce on average (the upper figure), while the averaged total estimation errors increase (the lower figure).

scheduling policy, and then designed the scheduler's parameters by solving a DP problem. The computational overhead of the proposed design method is less sensitive to the system dimension compared with that of existing algorithms in the literature.

Appendix. Proof of Lemmas 4, 5 and 2

Define $\tilde{z}_k = \hat{x}_k^s - A\hat{x}_{k-1}^s$. Evidently, we have $z_k = \sum_{j=\mathbb{N}_{k-1}+1}^{k} A^{k-j}\tilde{z}_j$. Some properties of $\tilde{z}_k$, which are preliminary for proving Lemma 4, are summarized in the following lemma. Since the lemma can be readily established from Shi, Johansson, and Qiu (2011, Lemma 2.2), The proof is omitted.

Lemma 7. *The following statements on $\tilde{z}_k$ hold:* (i). *$\tilde{z}_k$ is zero-mean Gaussian and $\mathbb{E}[\tilde{z}_k\tilde{z}_k'] = h(P_{k-1}^s) - P_k^s$;* (ii). *$\tilde{z}_1, \ldots, \tilde{z}_k, \ldots$ are independent.*

Proof of Lemma 4. First we shall prove the Gaussianity of $p_{\hat{\Theta}}(z_k, z|\mathcal{I}_{k-1}^{\mathrm{E}})$ by induction. For $k = 1$, we have $z_1 = \hat{x}_1^s - A\hat{x}_0^s$ and $\tilde{z}_1 \sim \mathcal{N}(0, \Sigma_1)$, where $\Sigma_1 = h(P_0^s) - P_1^s$ by Lemma 7. We assume $p_{\hat{\Theta}}(z_k|\mathcal{I}_{k-1}^{\mathrm{E}}) \sim \mathcal{N}(0, \Sigma_k)$ for some k. Define $\Omega_k = \{\Sigma_k^{1/2}\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}$. Then, with respect to the Lebesgue measure on Ω_k, we have

$$p_{\hat{\Theta}}(z_k, z|\mathcal{I}_{k-1}^{\mathrm{E}}) = \frac{1}{(2\pi)^{r_k/2}(\det \Sigma_k)^{1/2}} \exp\left(-\frac{1}{2}z'\Sigma_k^{-1}z\right),$$

where $r_k \triangleq \mathrm{rank}(\Sigma_k)$. Considering the triggering condition in (7), we have

$$\begin{aligned}\mathbb{P}\left(\gamma_k = 0|\mathcal{I}_{k-1}^{\mathrm{E}}\right) &= \int_{\Omega_k} \exp\left(-\frac{1}{2}z'\Gamma_k^{-1}z\right) p_{\hat{\Theta}}(z_k, z|\mathcal{I}_{k-1}^{\mathrm{E}})\mathrm{d}z \\ &= \int_{\Omega_k} \exp\left(-\frac{1}{2}z'(\Psi_k^{-1} - \Sigma_k^{-1})z\right) p_{\hat{\Theta}}(z_k, z|\mathcal{I}_{k-1}^{\mathrm{E}})\mathrm{d}z \\ &= \int_{\Omega_k} \frac{1}{(2\pi)^{r_k/2}(\det\Sigma_k)^{1/2}} \exp\left(-\frac{1}{2}z'\Psi_k^{-1}z\right) \mathrm{d}z \\ &= \det(\Psi_k\Sigma_k^{-1})^{1/2}. \end{aligned} \tag{24}$$

If $\gamma_k = 0$,

$$\begin{aligned} p_{\hat{\Theta}}\left(z_k, z|\mathcal{I}_{k-1}^{\mathrm{E}}, \gamma_k = 0\right) &= \frac{\mathbb{P}\left(\gamma_k = 0|z_k = z\right) p_{\hat{\Theta}}\left(z_k, z|\mathcal{I}_{k-1}^{\mathrm{E}}\right)}{\mathbb{P}\left(\gamma_k = 0|\mathcal{I}_{k-1}^{\mathrm{E}}\right)} \\ &= \frac{1}{(2\pi)^{r_k/2}(\det \Psi_k)^{1/2}} \exp\left(-\frac{1}{2}z'\Psi_k^{-1}z\right), \end{aligned} \tag{25}$$

i.e., $p_{\hat{\Theta}}\left(z_k, z|\mathcal{I}_{k-1}^{\mathrm{E}}, \gamma_k = 0\right) \sim \mathcal{N}(0, \Psi_k)$. Since $z_{k+1} = Az_k + \tilde{z}_{k+1}$, by Lemma 7,

$$p_{\hat{\Theta}}(z_{k+1}, z|\mathcal{I}_{k-1}^{\mathrm{E}}, \gamma_k = 0) \sim \mathcal{N}(0, \Sigma_{k+1}),$$

where $\Sigma_{k+1} = A\Psi_kA' + h(P_k^s) - P_{k+1}^s$. Otherwise, if $\gamma_k = 1$, $\hat{x}_k^s$ is sent to the estimator. Consequently, z_k conditioned on $\mathcal{I}_k^{\mathrm{E}}$ is deterministic and $z_{k+1} = \hat{x}_{k+1}^s - A\hat{x}_k^s = \tilde{z}_{k+1}$. By Lemma 7, $p_{\hat{\Theta}}(z_{k+1}, z|\mathcal{I}_{k-1}^{\mathrm{E}}, \gamma_k = 1) \sim \mathcal{N}(0, \Sigma_{k+1})$, where $\Sigma_{k+1} = h(P_k^s) - P_{k+1}^s$.

From optimal filtering theory, $x_k - \hat{x}_k^s$ is orthogonal to z_k. Since $x_k - \hat{x}_k^s$ and z_k are jointly Gaussian and $x_k - \hat{x}_k^s \sim \mathcal{N}(0, P_k^s)$, we reach the conclusion. ■

Proof of Lemma 5. The result is immediately from (24). ■

Proof of Lemma 2. Letting $z_k = \mathbf{y}^{(1)}$, $\Sigma_k = \sigma_{\mathbf{y}}^{(1)}$, $\Gamma_k = \Gamma^{(1)}$, $\Psi_k = ((\sigma_{\mathbf{y}}^{(1)})^{-1} + (\Gamma^{(1)})^{-1})^{-1} \triangleq \psi_1$ and $\mathcal{I}_{k-1}^{\mathrm{E}} = \emptyset$ in (24) and (25), then $p_{\hat{\theta}}(\mathbf{y}^{(1)}, y|\gamma^{(1)} = 0) \sim \mathcal{N}(0, \psi_1)$ follows from (25) and $\mathbb{P}_{\hat{\theta}}(\gamma^{(1)} = 0) = (\psi_1(\sigma_{\mathbf{y}}^{(1)})^{-1})^{1/2}$ follows from (15). Noting $\mathbb{P}_{\theta}(\gamma^{(1)} = 0) = 1 - 2q(\delta_1)$, we have $\psi_1 = (1 - 2q(\delta_1))^2\sigma_{\mathbf{y}}^{(1)}$. From optimal filtering theory, $\hat{\mathbf{x}}^{(1)}(\theta) = \frac{\sigma_{\mathbf{x}}}{\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)}}\mathbf{y}^{(1)}$, $\mathbf{x} - \hat{\mathbf{x}}^{(1)}(\theta)$ is orthogonal to $\mathbf{y}^{(1)}$ and $\mathbf{x} - \hat{\mathbf{x}}^{(1)}(\theta) \sim \mathcal{N}(0, \sigma_{\mathbf{x}} - \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)}})$. Then, we obtain $p_{\theta}(\mathbf{x}, x|\gamma^{(1)} = 0) \sim \mathcal{N}(0, \sigma)$ with $\sigma = \sigma_{\mathbf{x}} - \frac{4q(\delta_1)(1-q(\delta_1))\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}} + \sigma_{\mathbf{v}}^{(1)}}$, which completes the proof.

References

Anderson, B., & Moore, J. (1979). *Optimal filtering*. Prentice Hall.

Battistelli, G., Benavoli, A., & Chisci, L. (2012). Data-driven communication for state estimation with sensor networks. *Automatica*, *48*(5), 926–935.

Cheung, H. Y., Kwok, T. C., & Lau, L. C. (2013). Fast matrix rank algorithms and applications. *Journal of the ACM*, *60*(5), 31.

Chow, C.-S., & Tsitsiklis, J. N. (1989). The complexity of dynamic programming. *Journal of Complexity*, *5*(4), 466–488.

Gatsis, K., Ribeiro, A., & Pappas, G.J. (2013). Optimal power management in wireless control systems. In *Proc. American control conference* (pp. 1562–1569).

Han, D., Cheng, P., Chen, J., & Shi, L. (2013). An online sensor power schedule for remote state estimation with communication energy constraint. *IEEE Transactions on Automatic Control*, *59*(7), 1942–1949.

Han, D., Mo, Y., Wu, J., Weerakkody, S., Sinopoli, B., & Shi, L. (2015). Stochastic event-triggered sensor schedule for remote state estimation. *IEEE Transactions on Automatic Control*, *60*(10), 2661–2675.

Hovareshti, P., Gupta, V., & Baras, J. (2007). Sensor scheduling using smart sensors. In *Proc. IEEE conference on decision and control* (pp. 494–499).

Imer, O.C., & Başar, T. (2005). Optimal estimation with limited measurements. In *Proc. IEEE conference on decision and control* (pp. 1029–1034).

Li, L., Lemmon, M., & Wang, X. (2010). Event-triggered state estimation in vector linear processes. In *Proc. American control conference* (pp. 2138–2143).

Lipsa, G., & Martins, N. (2011). Remote state estimation with communication costs for first-order LTI systems. *IEEE Transactions on Automatic Control*, *56*(9), 2013–2025.

Molin, A. (2014). *Optimal event-triggered control with communication constraints*. (Ph.D. thesis), München: Technische Universität München, Diss..

Puterman, M. L. (2009). *Markov decision processes: discrete stochastic dynamic programming. Vol. 414*. Wiley.com.

Shi, D., Chen, T., & Darouach, M. (2016). Event-based state estimation of linear dynamic systems with unknown exogenous inputs. *Automatica*, *69*, 275–288.

Shi, D., Chen, T., & Shi, L. (2014). An event-triggered approach to state estimation with multiple point-and set-valued measurements. *Automatica*, *50*(6), 1641–1648.

Shi, D., Chen, T., & Shi, L. (2015). On set-valued Kalman filtering and its application to event-based state estimation. *IEEE Transactions on Automatic Control*, *60*(5), 1275–1290.

Shi, D., Elliott, R. J., & Chen, T. (2016). Event-based state estimation of discrete-state hidden Markov models. *Automatica*, *65*, 12–26.

Shi, L., Johansson, K.H., & Qiu, L. (2011). Time and event-based sensor scheduling for networks with limited communication resources. In *Proc. IFAC world congress* (pp. 13263–13268).

Sijs, J., & Lazar, M. (2012). Event based state estimation with time synchronous updates. *IEEE Transactions on Automatic Control*, *57*(10), 2650–2655.

Trimpe, S., & D'Andrea, R. (2014). Event-based state estimation with variance-based triggering. *IEEE Transactions on Automatic Control*, *59*(12), 3266–3281.

Tsitsiklis, J. N. (2007). Computational complexity in Markov decision theory. *HERMIS-An International Journal of Computer Mathematics and its Applications*, *9*, 45–54.

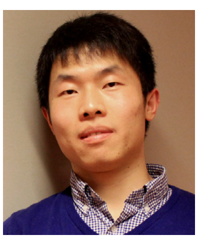
Wu, J., Jia, Q.-S., Johansson, K. H., & Shi, L. (2013). Event-based sensor data scheduling: Trade-off between communication rate and estimation quality. *IEEE Transactions on Automatic Control*, *58*(4), 1041–1046.

Wu, J., Johansson, K., & Shi, L. (2014). A stochastic online sensor scheduler for remote state estimation with time-out condition. *IEEE Transactions on Automatic Control*, *59*(11), 3110–3116.

Xu, Y., & Hespanha, J. (2005). Estimation under uncontrolled and controlled communications in networked control systems. In: *Proc. IEEE conference on decision and control and european control conference* (pp. 842–847).

Yang, C., & Shi, L. (2011). Deterministic sensor data scheduling under limited communication resource. *IEEE Transactions on Signal Processing*, *59*(10), 5050–5056.

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 the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology, Hong Kong. He 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 wirelesssensor networks, multi-agent systems. He received the Guan Zhao-Zhi Best Paper Award at the 34th Chinese Control Conference in 2015.



Xiaoqiang Ren received the B.E. degree in the Department of Control Science and Engineering from Zhejiang University, Hangzhou, China, in 2012. He is currently working towards the Ph.D. degree in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology. He was a visiting student in ACCESS Linnaeus Centre at the KTH Royal Institute of Technology, Sweden in 2015. His research interests include sequential detection, security of cyber–physical system and networked control and estimation.

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 with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA, USA in 2014. He is currently a Research Fellow at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include cyber–physical security, wireless sensor scheduling and event-based state estimation.

Dawei Shi received his B.Eng. degree in Electrical Engineering and its Automation from the Beijing Institute of Technology in 2008. He received his Ph.D. degree in Control Systems from the University of Alberta in 2014. Since December 2014, he has been appointed as an Associate Professor at the School of Automation, Beijing Institute of Technology, China. His research interests include event-based control and estimation, robust model predictive control and tuning, and wireless sensor networks. He is a reviewer for a number of international journals, including IEEE Transactions on Automatic Control, Automatica, and Systems & Control Letters. In 2009, he received the Best Student Paper Award in IEEE International Conference on Automation and Logistics.

Ling Shi received the B.S. degree in Electrical and Electronic Engineering from Hong Kong University of Science and Technology, Hong Kong, in 2002 and the Ph.D. degree in Control and Dynamical Systems from California Institute of Technology, 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.