A Tutorial on Remote State Estimation

Mohammed Shalaby, 260607172

***Abstract*—The following paper provides an overview of three of the keymark papers in remote estimation, with a focus on their proofs. In addition, we run simulations to compare the performance of two of the strategies.**

I. Introduction

The recent advancement in wireless communication have motivated the study of networked control systems, which have various applications. To be able to achieve a satisfactory performance, a networked control system has to find a communication strategy between the sensor and the controller that attempts to communicate information only when the information is needed, such that the controller can achieve a stable performance. To do so, the controller requires a state estimator, and the problem then involves finding a communication strategy and an estimator that reduces the communication burden while ensuring a reasonable estimate is always available to the controller.

Reasons behind limiting the communication between the sensor and the estimator include the lack of power available for the agents and a limited bandwidth for the communication scheme, and thus by limiting the number of transmissions made between each sensor and estimator, priority is given to more meaningful transmissions. Meanwhile, the estimation quality can be measured using the covariance of the error between the true value and the estimated value.

Most of the literature focuses on obtaining threshold-type transmission policies that are optimal in some sense. [1] address finding the optimal communication policy by formulating it as a long term average cost optimization problem, and solving it using dynamic programming. They show that for discrete-time LTI systems with fixed network delays, the best communication policy is deterministic and the optimal rules are provided. In [2], the users formulate a problem with perfect-state feedback and show that the optimal communication policy is of the symmetric threshold type, and that the optimal estimator is Kalman-like. In [3] and [4], they show that for the case with packet-drop, the optimal policy is of the threshold-type and the optimal estimator is similar to a Kalman filter. The work in [5] uses a variance-based communication scheme, and they show that the covariance recursion asymptotically converges to a periodic one. Most recently, [6] have addressed the remote estimation problem in which the sensor observes a Markovian state and chooses the power level to transmit it over a time-varying packet-drop channel. They show that, yet again, the best communication policy is of the symmetric threshold type, and is monotone, while the optimal estimation strategy is similar to a Kalman filter.

The main disadvantage of such deterministic communication policies is that for cases with partial-state feedback, event-based scheduling results in a truncated Gaussian distribution which deems the MMSE estimator intractable. In [7], the authors developed an approximate MMSE that maintains the deterministic communication policy. In [8] and [9], the communication policy is replaced with a suboptimal stochastic scheduler which then allows the authors to derive a tractable exact MMSE estimator. In [10], the problem is looked at from a different approach, where instead of attempting to find the optimal communication policy, they find the optimal value to be transmitted by the sensor for different set ups.

This paper is split as follows. We start by formulating the remote estimation problem in Section II, where we assume an ideal communication channel. Section III then addresses the deterministic communication policies, with Section III-A addressing the work in [2] and Section III-B addressing the work in [7]. We then address the stochastic communication policies in Section IV, with a main focus on the work in [8]. Simulations are then run and the results compared for the different strategies in Section V, and Section VI discusses a few final remarks and a word on future work.

Notation: A bold variable $\mathbf{x}$ denotes a vector, x denotes a scalar, and a capital bold variable $\mathbf{X}$ denotes a matrix. $\mathbf{X} > 0$ ($\mathbf{X} \geq 0$) means $\mathbf{X}$ is positive (semi-)definite, and $\mathbf{X} \geq \mathbf{Y}$ means $\mathbf{X} - \mathbf{Y} \geq 0$. $\mathbb{S}^n_+$ ($\mathbb{S}^n_{++}$) denotes the set of $n \times n$ positive (semi)-definite matrices. $f_x(y)$ denotes the probability density function (PDF) of random variable x at the realization y, and $f_{x|z}(y|z)$ denotes the conditional PDF of x given z. Meanwhile, $p_x(y)$ denotes the probability mass function (PMF) of random variable x at the realization y. $\mathbb{E}^{\Pi}_{f_{x|z}}(x|z)$ denotes the conditional expectation w.r.t. the PDF $f_{x|z}$. given a policy Π. $\mathbb{P}[\cdot|\cdot]$ is reserved for conditional probabilities. $\|\cdot\|_\infty$ denotes the Hölder infinity-norm of a vector, i.e., $\|\boldsymbol{v}\|_\infty = \max\{|v_1|, |v_2|, \ldots\}$. $\mathrm{Tr}(\cdot)$ denotes the trace operator, $*$ denotes convolution, and $\circ$ is the function composition. We also denote $\{v_k\}_{k=0}^T$ as $v_{0:T}$. Whenever the intended meaning is clear, some of the notation will be dropped for conciseness.

II. Problem Formulation

Consider a linear, time-invariant system in discrete-time

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{w}_k, \tag{1}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k, \quad \forall k \geq 0, \tag{2}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector, $\mathbf{y}_k \in \mathbb{R}^m$ is the sensor measurement, $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \in \mathbb{R}^n$ is the process noise, and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \in \mathbb{R}^m$ is the measurement noise. Both $\mathbf{w}_k$ and

$\mathbf{v}_k$ are assumed to be mutually uncorrelated, and the matrices $\mathbf{Q} \geq 0$ are $\mathbf{R} > 0$ are covariance matrices.

The initial state $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_0)$ is unknown, and the covariance matrix $\mathbf{P}_0 = \mathbb{E}[\mathbf{x}_0 \mathbf{x}_0^\mathsf{T}] \geq 0$ dictating the uncertainty of the initial state is assumed to be uncorrelated with $\mathbf{w}_k$ and $\mathbf{v}_k$ for all $k \geq 0$. The system is assumed to be observable and controllable at all times.

Unlike common estimation problems, remote state estimation addresses the challenges introduced when a sensor is connected to the estimator through a costly communication scheme. At every time step, the sensor has two options: transmit the most recent information to the estimator or not. Let $\gamma_k \in \{0, 1\}$ denote the communication decision made by the scheduler, with $\gamma_k = 1$ denoting that the scheduler decides to transmit the message.

At time step k, the information set available to the sensor is given by

$$\mathcal{I}_k^\mathrm{s} \triangleq \{\mathbf{y}_1, \ldots, \mathbf{y}_k\} \cup \{\gamma_1, \ldots, \gamma_k\}, \quad \text{with } \mathcal{I}_{-1}^\mathrm{s} = \emptyset.$$

By defining the transmitted signal to be

$$\mathbf{s}_k \triangleq \mathbf{q}_k(\mathcal{I}_k^\mathrm{s}),$$

the information set available to the estimator at time step k is

$$\mathcal{I}_k^\mathrm{e} \triangleq \{\gamma_1 \mathbf{s}_1, \ldots, \gamma_k \mathbf{s}_k\} \cup \{\gamma_1, \ldots, \gamma_k\}, \quad \text{with } \mathcal{I}_{-1}^\mathrm{e} = \emptyset.$$

The communication scheduling policy applied by the sensor at time k is then defined as

$$\gamma_k = c_k(\mathbf{y}_k, \mathcal{I}_{k-1}^\mathrm{s}),$$

where c_k is assumed to be a measurable mapping $\forall k$. A finite-horizon sensor communication policy Θ is accordingly defined as a sequence of c_k's,

$$\Theta \triangleq \{c_1, \ldots, c_T\},$$

where T denotes the horizon. The estimator is then designed such that it has knowledge of the communication policy Θ being used by the scheduler, and computes its own estimate as per

$$\hat{\mathbf{x}}_k = \boldsymbol{e}_k(\mathcal{I}_k^\mathrm{e}),$$

where $\boldsymbol{e}$ is assumed to be a measurable mapping, $\forall k$. Similarly, a finite-horizon estimator Ξ is defined as

$$\Xi \triangleq \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_T\}.$$

The penalty associated with the total expected estimation errors of the remote estimator within a horizon T is given by

$$J_\mathrm{e}(\Theta, \Xi), \triangleq \mathbb{E}\left[\sum_{k=1}^{T} \mathrm{Tr}\left(\mathbb{E}^{\Theta,\Xi}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^\mathsf{T} | \mathcal{I}_k^\mathrm{e}\right]\right)\right],$$

and the communication cost incurred due to the sensor-to-estimator communication times within the horizon T is given by

$$J_\mathrm{c}(\Theta) \triangleq \mathbb{E}\left[\sum_{k=1}^{T} \mathbb{E}^{\Theta}\left[\gamma_k | \mathcal{I}_{k-1}^\mathrm{e}\right]\right].$$

Therefore, the optimization problem associated with minimizing the total cost of the system is given by

$$J^*(\lambda, \Theta, \Xi) \triangleq \min_{(\Theta,\Xi) \in \mathscr{C} \times \mathscr{E}} \{J_e(\Theta, \Xi) + \lambda J_c(\Theta)\}, \tag{3}$$

where λ is a Lagrange coefficient, and can be thought of as a weighting parameter of the cost associated with the communication channel.

In addition, we define

$$\begin{aligned} \check{\mathbf{x}}_k &\triangleq \mathbb{E}[\mathbf{x}_k | \mathcal{I}_{k-1}^\mathrm{e}], \\ \check{\mathbf{e}}_k &\triangleq \mathbf{x}_k - \check{\mathbf{x}}_k, \\ \check{\mathbf{P}}_k &\triangleq \mathbb{E}[\check{\mathbf{e}}_k \check{\mathbf{e}}_k^\mathsf{T} | \mathcal{I}_{k-1}^\mathrm{e}], \end{aligned} \tag{4}$$

and

$$\begin{aligned} \hat{\mathbf{x}}_k &\triangleq \mathbb{E}[\mathbf{x}_k | \mathcal{I}_k^\mathrm{e}], \\ \hat{\mathbf{e}}_k &\triangleq \mathbf{x}_k - \hat{\mathbf{x}}_k, \\ \mathbf{P}_k &\triangleq \mathbb{E}[\hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^\mathsf{T} | \mathcal{I}_k^\mathrm{e}]. \end{aligned} \tag{5}$$

The estimates $\check{\mathbf{x}}_k$ and $\hat{\mathbf{x}}_k$ are called the *a priori* and *a posteriori* MMSE estimate, respectively. In addition, define the measurement innovation $\mathbf{z}_k$ as

$$\mathbf{z}_k \triangleq \mathbf{y}_k - \mathbb{E}[\mathbf{y}_k | \mathcal{I}_{k-1}^\mathrm{e}]. \tag{6}$$

III. Deterministic Communication Policy

The majority of this report attends to deterministic communication policies, which are policies of the form

$$\gamma_k = \begin{cases} 0, & \text{if } c_k(\mathbf{y}_k, \mathcal{I}_{k-1}^\mathrm{s}) \leq \delta_k, \\ 1, & \text{otherwise.} \end{cases}$$

We first show that optimal communication policies are of this form, and then address the derivation of an optimal estimator.

A. Perfect State Feedback

We first address the case of a scalar system with perfect state feedback [2], i.e., (1) and (2) become

$$x_{k+1} = a x_k + w_k, \tag{7}$$

$$y_k = x_k, \quad \forall k > 0. \tag{8}$$

Consider a free symbol denoted $\mathfrak{E}$ and a pre-processor $\mathcal{P}_{0,T} : (y_0, \ldots, y_k) \to s_k$ defined for $k \in \{0, \ldots, T\}$, and $s_k \in \mathbb{R} \cup \{\mathfrak{E}\}$. Note that the free symbol $\mathfrak{E}$ represents the decision of the pre-processor not to transmit the measurement to the estimator, and thus has a corresponding communication cost of 0. We assume that $s_0 = y_0$.

Given the pre-processor, consider optimal estimators whose output at time k is denoted as $\hat{x}_k$ and is expressed as

$$\hat{x}_k = \begin{cases} \mathbb{E}\left[x_k | \mathcal{I}_k^\mathrm{e}\right], & \text{if } k \geq 1, \\ x_0, & \text{if } k = 0. \end{cases}$$

The notation $\mathcal{E}(\mathcal{P}_{0,T})$ is used to denote the *optimal estimator* associated with a given pre-processor $\mathcal{P}_{0,T}$.

Given a discounting factor $d \in (0,1]$ and a positive real constant c, the cost function to be minimized is then given as

$$J_{0,T}(\mathcal{P}_{0,T}) \triangleq \sum_{k=1}^{T} d^{k-1}\mathbb{E}[(x_k - \hat{x}_k)^2 + cr_k], \tag{9}$$

where r_k is defined as per the indicator function

$$r_k \triangleq \mathbb{I}(s_k \neq \mathfrak{E}) = \begin{cases} 0, & \text{if } s_k = \mathfrak{E}, \\ 1, & \text{otherwise.} \end{cases}$$

The goal is then to solve the optimization problem given by

$$\mathcal{P}^*_{0,T} = \arg\min_{\mathcal{P}^*_{0,T}\in\mathscr{P}} J_{0,T}(\mathcal{P}_{0,T}), \tag{10}$$

where $\mathcal{P}^*_{0,T}$ is the optimal pre-processor policy, and $\mathscr{P}$ is the set of all admissible policies. Note that this represents a discounted form of (3).

The authors start with a particular choice of estimator and pre-processor, which are denoted as Kalman-like and symmetric threshold policy, respectively. This leads to Theorem III.1, which states that these operators are optimal for (10), as will be proven towards the end of this section.

Definition III.1. (Kalman-like estimator) Given the process model defined in (7) and a pre-processor $\mathcal{P}_{0,T}$, define the map $H : (s_0, \dots, s_k) \to \eta_k$, for $k \in \{0, \dots, T\}$, where η_k is computed using

$$\eta_0 = x_0,$$
$$\eta_k = \begin{cases} a\eta_{k-1}, & \text{if } s_k = \mathfrak{E}, \\ s_k, & \text{otherwise,} \end{cases} \qquad \forall k \geq 1.$$

Definition III.2. (Admissible Pre-processor) Let a horizon $T > 0$ and a pre-processor policy $\mathcal{P}_{0,T}$ be given. The pre-processor $\mathcal{P}_{0,T}$ is admissible if there exist maps $\mathcal{P}_{m,T} : (x_m, \dots, x_k) \to s_k$ (note that we can replace y_i with x_i due to (8)), with $0 \leq m \leq T$ and $k \geq m$, such that $\mathcal{P}_{0,T}$ can be specified recursively as follows:

Description of the Algorithm for $\mathcal{P}_{m,T}$.

- (Initial Step) Set $k = m$, $r_m = 1$, and transmit the current state, i.e., $s_m = x_m$.
- (Step A) Increase the counter k by one. If $k > T$ holds then terminate, otherwise execute Step B.
- (Step B) Obtain the pre-processor output at time k via

  $$s_k = \mathcal{P}_{m,T}(x_m, \dots, x_k).$$

  If $s_k = \mathfrak{E}$ then set $r_k = 0$ and go back to Step A. If $s_k \neq \mathfrak{E}$ then execute algorithm $\mathcal{P}_{k,T}$.

End of the description of the Algorithm for $\mathcal{P}_{m,T}$.

Basically, you start at time step 0 and transmit the measurement, then you consider $\mathcal{P}_{0,1}$. If you have to transmit, then you set $r_1 = 1$ and the time step 1 is your new 0, and you execute $\mathcal{P}_{1,T}$ the same way recursively. Otherwise, you set $r_1 = 0$, and look at $\mathcal{P}_{0,2}$, then $\mathcal{P}_{0,3}$, until you a find a new point where you transmit the message, and then you can restart from there. The class of all admissible pre-processors is denoted as $\mathbb{P}_T$.

Remark III.1. Given a positive time-horizon T, there is no loss of generality in constraining our search for an optimal pre-processor to the set $\mathbb{P}_T$ (i.e., there always exists an admissible pre-processor that is an optimal solution). In order to justify this, consider that an optimal pre-processor policy $\mathcal{P}^*_{0,T}$ is given. If a transmission takes place at some time m (i.e., $r_m = 1$), then the optimal output at the pre-processor is $s_k = x_k$. In fact, given that a real number is transmitted, the choice $s_k = x_k$ must be optimal because it leads to no estimation error. Hence, given that $r_m = 1$, by Markovianity we conclude that the current and future output produced by the pre-processor $s_{m:T}$ will not depend on the state x_k for times k prior to m. Consequently, $\mathcal{P}^*_{0,T}$ is admissible.

The prediction error can be simply denoted as

$$\check{e}_k \triangleq x_k - a\eta_{k-1}.$$

Using (7) and Definition III.1,

$$\check{e}_0 = 0, \tag{11}$$
$$\begin{aligned}\check{e}_{k+1} &= x_{k+1} - a\eta_k \\ &= ax_k + w_k - a\begin{cases} a\eta_{k-1}, & \text{if } r_k = 0, \\ x_k, & \text{if } r_k = 1, \end{cases} \\ &= \begin{cases} a(x_k - a\eta_{k-1}) + w_k, & \text{if } r_k = 0, \\ w_k, & \text{if } r_k = 1, \end{cases} \\ &= \begin{cases} a\check{e}_k + w_k, & \text{if } r_k = 0, \\ w_k, & \text{if } r_k = 1. \end{cases}\end{aligned} \tag{12}$$

In addition, the cost function in (9) in terms of $\check{e}_k$ is

$$J_{0,T}(\mathcal{P}_{0,T}) = \sum_{k=1}^{T} d^{k-1}\mathbb{E}[(\check{e}_k - \mathbb{E}[\check{e}_k | \mathcal{I}^{\mathrm{e}}_k])^2 + cr_k]. \tag{13}$$

A key fact here is that $\mathbb{E}[\check{e}_k | \mathcal{I}^{\mathrm{e}}_k] = \mathbb{E}[x_k | \mathcal{I}^{\mathrm{e}}_k] - a\eta_{k-1}$ holds, since,

$$\begin{aligned}\mathbb{E}[\check{e}_k | \mathcal{I}^{\mathrm{e}}_k] &= \mathbb{E}[x_k - a\eta_{k-1} | \mathcal{I}^{\mathrm{e}}_k] \\ &= \mathbb{E}[x_k | \mathcal{I}^{\mathrm{e}}_k] - \mathbb{E}[a\eta_{k-1} | \mathcal{I}^{\mathrm{e}}_k] \\ &= \mathbb{E}[x_k | \mathcal{I}^{\mathrm{e}}_k] - a\eta_{k-1}.\end{aligned}$$

This leads to the validity of the identity $\check{e}_k - \mathbb{E}[\check{e}_k | \mathcal{I}^{\mathrm{e}}_k] = x_k - \mathbb{E}[x_k | \mathcal{I}^{\mathrm{e}}_k]$.

Definition III.3. (Symmetric threshold policy) Given a positive integer horizon T and an arbitrary sequence of positive real numbers (thresholds) $\tau_{1:T}$, for each $m \in \{0, \dots, T\}$, we define the following algorithm for $k \geq m$, which is denoted as $\mathcal{S}_{m,T}$,

1) (Initial step) Set $k = m$, $r_m = 1$, and transmit the current state, i.e., $s_m = x_m$ or equivalently set $\check{e}_m = 0$.

2) (Step A) Increase the time counter k by one. If $k > T$ holds then terminate; otherwise, execute Step B.

3) (Step B) If $|\check{e}_k| < \tau_k$ holds, then set $r_k = 0$, transmit the free symbol, i.e., $s_k = \mathfrak{E}$, and return to Step A. If $|\check{e}_k| \geq \tau_k$ holds, then set $m = k$ and execute $\mathcal{S}_{m,T}$.

This algorithm $\mathcal{S}_{0,T}$ is denoted as a symmetric threshold pre-processor.

Remark III.2. Note that this is possible at the sensor since it is aware of both the true state and all the inputs to the estimator, thus can compute the prediction error $\check{e}_k$.

The main result of the paper then follows.

Theorem III.1. There exists a sequence of positive real numbers $\tau^*_{1:T}$, such that the corresponding symmetric threshold policy $\mathcal{S}^*_{0,T}$ is an optimal solution to (10) and the corresponding optimal estimator $\mathcal{E}(\mathcal{S}^*_{0,T})$ is H.

We start by defining the following class of path-dependent pre-processor policies, which is an extension so as to allow time-varying thresholds that depend on past decisions.

Definition III.4. (Path-dependent symmetric threshold policy) Given a horizon T, consider that a sequence of (threshold) functions $\mathcal{T} \triangleq \{\mathcal{T}_{m,k} | m \leq k \leq T, 1 \leq m \leq T\}$, with $\mathcal{T}_{m,k} : \{0,1\}^{m-k} \to \mathbb{R}$, is given. For every m in the set $\{1, \ldots, T\}$, we define the following algorithm, which we denote as $\mathcal{D}_{m,T}$:

Description of the Algorithm for $\mathcal{D}_{m,T}$.

- (Initial Step) Set $k = m$, $r_m = 1$, and transmit the current state, i.e., $s_m = x_m$ or equivalently set $\check{e}_m = 0$.
- (Step A) Increase the counter k by one. If $k > T$ holds then terminate, otherwise execute Step B.
- (Step B) If $|\check{e}_k| < \mathcal{T}_{m,k}(r_m, \ldots, r_{k-1})$ holds then set $r_k = 0$, transmit the free symbol (i.e., $s_k = \mathfrak{E}$), and return to Step A. If $|\check{e}_k| \geq \mathcal{T}_{m,k}(r_m, \ldots, r_{k-1})$ holds then execute $\mathcal{D}_{k,T}$.

End of the description of the Algorithm for $\mathcal{D}_{m,T}$.

We use $\mathbb{D}_{0,T}$ to denote the entire class of path-dependent symmetric threshold pre-processors with time horizon T.

Proposition III.1. Let $\mathcal{D}_{0,T}$ be a preselected path-dependent symmetric threshold policy, it holds that the optimal estimator $\mathcal{E}(\mathcal{D}_{0,T})$ is H.

Proof. We have that $\hat{e}_k$ as defined in (5) can be expressed as

$$\begin{aligned}
\hat{e}_0 &= 0, \\
\hat{e}_{k+1} &= x_{k+1} - \eta_{k+1} \\
&= x_{k+1} - \begin{cases} a\eta_k, & \text{if } r_k = 0, \\ x_{k+1}, & \text{if } r_k = 1, \end{cases} \\
&= \begin{cases} a(x_k - \eta_k) + w_k, & \text{if } r_k = 0, \\ 0, & \text{if } r_k = 1, \end{cases} \\
&= \begin{cases} a\eta_k + w_k, & \text{if } r_k = 0, \\ 0, & \text{if } r_k = 1. \end{cases}
\end{aligned}$$

We then show that the probability density function of $\hat{e}_k$ is even by induction. Since $\hat{e}_0 = 0$, we assume the symmetry of the PDF of $\hat{e}_k$ as our induction basis, then the PDF of $\hat{e}_{k+1}$ is symmetric as it is either 0, or the sum of two zero-mean Gaussian random variables.

Consequently, under symmetric path-dependent threshold policies, the PDF of $\hat{e}_k$ given the past and current observations $\{v_t\}_{t=0}^k$ is even. Therefore,

$$\mathbb{E}[\hat{e}_k | \mathcal{I}^{\mathrm{e}}_k] = 0,$$

which implies that

$$\hat{x} = \mathbb{E}[x_k | \mathcal{I}^{\mathrm{e}}_k] = \eta_k.$$

□

Remark III.3. By doing a similar proof by induction as was used to show the symmetry of $\hat{e}_k$, we can show that given a symmetric path-dependent threshold pre-processor $\mathcal{D}_{0,T}$, we have that $\mathbb{E}[\check{e}_k \big| \mathcal{I}^{\mathrm{e}}_k] = 0$. Therefore, (13) becomes

$$J_{0,T}(\mathcal{D}_{0,T}) = \sum_{k=1}^{T} d^{k-1} \mathbb{E}[\check{e}_k^2 + c r_k], \quad \mathcal{D}_{0,T} \in \mathbb{D}_T. \tag{14}$$

The evolution of the prediction error defined in (12) can be seen as a Markov Decision Process (MDP) whose state and control are $\check{e}_k$ and r_k, respectively. Hence the minimization of (14) with respect to pre-processor policies $\mathcal{D}_{0,T}$ in the class $\mathbb{D}_k$ can be written as a dynamic program. To do so, we define the sequence of value functions $V_t : \mathbb{R} \to \mathbb{R}$, $t \in \{1, \ldots, T+1\}$ as observed by the pre-processor. Here t denotes the time at which decision was taken, and the argument of the function is the MDP state $\check{e}_t$.

Using dynamic programming, we set up the recursive equations

$$\begin{aligned}
V_{T+1}(\check{e}_{T+1}) &\triangleq 0, \\
V_t(\check{e}_t) &\triangleq \min_{r_t \in \{0,1\}} \mathcal{C}_t(\check{e}_t, r_t), \quad t \in \{1, \ldots, T\},
\end{aligned}$$

where $\mathcal{C} : \mathbb{R} \times \{0,1\} \to \mathbb{R}$ is defined as

$$\mathcal{C}_t(\check{e}_t, r_t) \triangleq \begin{cases} c + d\mathbb{E}[V_{t+1}(w_t)], & \text{if } r_t = 1, \\ \check{e}_t^2 + d\mathbb{E}[V_{t+1}(a\check{e}_t + w_t)], & \text{if } r_t = 0. \end{cases} \tag{15}$$

From (15), it immediately follows that an optimal decision policy r^*_t at any time t is given by

$$r^*_t = \begin{cases} 1, & \text{if } \mathcal{C}_t(\check{e}_t, 1) \leq \mathcal{C}_t(\check{e}_t, 0), \\ 1, & \text{if } \mathcal{C}_t(\check{e}_t, 0) < \mathcal{C}_t(\check{e}_t, 1). \end{cases} \tag{16}$$

Lemma III.1. Consider (10) with the additional constraint that the pre-processor must be of the symmetric path-dependent type $\mathbb{D}_t$. Then, there exists an optimal path-*independent* symmetric threshold policy $\mathcal{S}^*_{0,T}$, whose associated threshold selection $\tau^*_{1:T}$ is given by a solution to the following equation at every time t,

$$\mathcal{C}_t(\tau^*_t, 0) = \mathcal{C}_t(\tau^*_t, 1), \quad t \in \{1, \ldots, T\}. \tag{17}$$

Proof. From (16) and the Definition (III.3) of $\mathcal{S}_{0,T}$, it is sufficient to prove that there exist thresholds $\tau^*_{1:T}$ for which

$$|\check{e}_t| \geq \tau^*_t \iff \mathcal{C}_t(\check{e}_t, 1) \leq \mathcal{C}_t(\check{e}_t, 0)$$

holds $\forall t \in \{1, \ldots, T\}$.

Note that given (15), we have that $\mathcal{C}_t(\check{e}_t, 1) = \mathcal{C}_t(1)$, $\forall t \in \{1, \ldots, T\}$, and

$$\mathcal{C}_t(0, 0) < \mathcal{C}_t(\check{e}_t, 1)$$

for $\check{x}_t \in \mathbb{R}$, since $c > 0$.

We then show by induction that $\mathcal{C}_t(\check{e}_t, 0)$ is a continuous, even, quasi-convex and unbounded function of $\check{e}_t$, for every t in the set $\{1, \ldots, T\}$. Note that from (15), it is sufficient to show that $V_t(\check{e}_t)$ is even, quasi-convex, bounded, and continuous, of $\check{e}_t$ since $\check{e}_t^2$ is even, convex, unbounded and continuous function of $\check{e}_t$.

Since $V_{T+1}(\check{e}_{T+1}) = 0$, it satisfies all the properties by convention. Assuming that $V_{t+1}(\check{e}_{t+1})$ is even, convex, unbounded and continuous function of $\check{e}_t$, we have that $V_t(\check{e}_t)$ is even, convex, unbounded and continuous function of $\check{e}_t$. Therefore, by Lemma A.1 and by induction, we have that $\mathbb{E}[V_t(a\check{e}_{t-1} + w_{t-1})]$ is even, convex, unbounded and continuous function of $\check{e}_t$, $\forall t \in \{1, \ldots, T\}$.

Due to the independence of $\mathcal{C}_t(\check{e}_t, 1)$ from $\check{e}$, and the proven properties of $\mathcal{C}_t(\check{e}_t, 0)$, then $\exists u_t$ large enough such that

$$\mathcal{C}_t(\check{e}_t, 0) > \mathcal{C}_t(\check{e}_t, 1),$$
$$\mathcal{C}_t(-\check{e}_t, 0) > \mathcal{C}_t(-\check{e}_t, 1),$$

whenever $|\check{e}_t| > u_t$. Therefore, this implies that there exists a unique solution $\tau^*_{1:T}$ to (16). □

Let a pre-processor $\mathcal{P}_{0,T}$ be given. Following the path of the paper, we define the following notation for conciseness:

1) Define the conditional probability density function of $\check{e}_k$ given that only free symbols were transmitted up until time j as
$$\gamma_{k|j}(e) \triangleq f_{\check{e}_k|r_1=0,\ldots,r_j=0}(e), \quad e \in \mathbb{R}.$$
2) Define the probability that, under policy $\mathcal{P}_{0,T}$, only free symbols have been transmitted up until time k as
$$\varsigma_k \triangleq \begin{cases} \mathbb{P}(r_1 = 0, \ldots, r_k = 0), & \text{if } k \geq 1, \\ 1, & \text{if } k = 0. \end{cases}$$
3) Define the conditional probability that, under policy $\mathcal{P}_{0,T}$, the pre-processor transmits the free symbol at time k, given that only free symbols have been tranmitted up until time j as
$$\varsigma_{k|j} \triangleq \begin{cases} \mathbb{P}(r_k = 0 | r_1 = 0, \ldots, r_j = 0), & \text{if } k \geq 1, \\ \varsigma_1, & \text{if } k = 0. \end{cases}$$
4) Define the probability $\rho_k(e) : \mathbb{R} \to [0, 1]$ that, at time k, the free symbol is transmitted, given that $\check{e}_k = e$, where e is any real number, and the fact that only free symbols have been transmitted up until time $k-1$ as
$$\rho_k(e) \triangleq \mathbb{P}(r_k = 0 | \check{e}_k = e, r_1 = 0, \ldots, r_{k-1} = 0).$$

The probability density function of $a\check{e}_k$ given that no observation was received up until time k is denoted

$$\gamma^a_{k|j}(e) \triangleq f_{a\check{e}_k|r_1=0,\ldots,r_j=0}(e), \quad e \in \mathbb{R}.$$

If we denote the probability density function of w_k by $\mathcal{N}_{\sigma_w^2}$, then by (12) and Lemma B.1,

$$\gamma_{k|k-1}(e) = \gamma^a_{k-1|k-1}(e) * \mathcal{N}_{\sigma_w^2}. \tag{18}$$

Using Bayes' rule, we have that

$$\gamma_{k|k}(e) = \frac{\gamma_{k|k-1}(e)\rho_k(e)}{\varsigma_{k|k-1}}, \quad \varsigma_{k|k-1} \neq 0, k \geq 0. \tag{19}$$

Definition III.5. Given an admissible pre-processor $\mathcal{P}_{0,T}$ and an integer $m \in \{0, \ldots, T\}$, we adopt the following definition for the partial cost computed for the horizon $\{m+1, \ldots, T\}$ under the assumption that $r_m = 1$:

$$J_{m,T}(\mathcal{P}_{m,T}) \triangleq \begin{cases} \Sigma_{m+1}^T d^{k-m-1} \mathbb{E}\left[(\check{e}_k - \mathbb{E}[\check{e}_k | \mathcal{I}^{\mathrm{e}}_k])^2 + c r_k\right], & \text{if } m < T, \\ 0, & \text{if } m = T. \end{cases}$$

Proposition III.2. Given an arbitrarily selected admissible pre-processor $\mathcal{P}_{0,T}$, the finite horizon cost (13) can be expanded as

$$J_{0,T}(\mathcal{P}_{0,T}) = \sum_{k=1}^T d^{k-1} \Big(\mathbb{E}_{\gamma_{k|k}}\left[(\check{e}_k - \mathbb{E}[\check{e}_k | \mathcal{I}^{\mathrm{e}}_k])^2\right] \varsigma_k + (c + J_{k,T}(\mathcal{P}_{k,T}))\varsigma_{k-1}(1 - \varsigma_{k|k-1}) \Big).$$

Proof. The cost-to-go at every time-step is given by the cost incurred at that time-step plus all future costs, therefore (13) can be written as

$$\begin{aligned} &J_{0,T}(\mathcal{P}_{0,T}) \\ &= \sum_{k=1}^T d^{k-1} \Big(\mathbb{E}\left[\left(\check{e}_k - \mathbb{E}[\check{e}_k | \mathcal{I}^{\mathrm{e}}_k]\right)^2 | r_1 = 0, \ldots, r_k = 0 \right] \\ &\quad \times (\mathbb{E}[J_{k,T}(\mathcal{P}_{k,T}) | r_k = 1, r_1 = 0, \ldots, r_{k-1} = 0] + c) \\ &\quad \times \mathbb{P}(r_k = 1, r_1 = 0 \ldots, r_{k-1} = 0) \Big). \end{aligned}$$

By markovianity, we have that

$$\mathbb{E}[J_{k,T}(\mathcal{P}_{k,T}) | r_k = 1, r_1 = 0, \ldots, r_{k-1} = 0] = J_{k,T}(\mathcal{P}_{k,T}).$$

In addition,

$$\begin{aligned} &\mathbb{P}(r_k = 1, r_1, \ldots, r_{k-1} = 0) \\ &= \mathbb{P}(r_k = 1 | r_1 \ldots, r_{k-1} = 0)\mathbb{P}(r_1 = 0, \ldots, r_{k-1} = 0) \\ &= (1 - \mathbb{P}(r_k = 1 | r_1 \ldots, r_{k-1} = 0))\mathbb{P}(r_1 = 0, \ldots, r_{k-1} = 0) \\ &= \varsigma_{k-1}(1 - \varsigma_{k|k-1}), \end{aligned}$$

which concludes this proof. □

Definition III.6. The following is a convenient definition for the optimal cost:

$$J^*_{m,T} \triangleq \min_{\mathcal{P}_{m,T}\in\mathbb{P}_{m-T}} J_{m,T}(\mathcal{P}_{m,T}), \quad T\geq 1,$$

and $J^*_{m,0}=0$.

Corollary III.1. The following inequality holds for every admissible pre-processor $\mathcal{P}_{0,T}$:

$$J_{0,T}(\mathcal{P}_{0,T}) \geq \sum_{k=1}^{T} d^{k-1}\big(\mathbb{E}_{\gamma_{k|k}}\left[(\check{e}_k-\mathbb{E}[\check{e}_k|\mathcal{I}^{\mathrm{e}}_k])^2\right]\varsigma_k \\ + (c+J^*_{k,T}(\mathcal{P}_{k,T}))\varsigma_{k-1}(1-\varsigma_{k|k-1})\big).$$

Proof. This directly follows from Proposition III.2 and Definition III.6. □

Definition III.7. (Neat PDF) Let $f:\mathbb{R}\to\mathbb{R}$ be a probability density function. We say that f is neat if f is quasi-concave and there exists a real number b such that f is non-increasing on the interval $(-\infty,b)$ and non-decreasing on $[b,\infty)$.

We now have all preliminaries to prove Theorem III.1.

Proof. Proof of Theorem III.1

The strategy to prove Theorem III.1 is to show that for every admissible pre-processor policy $\mathcal{P}_{0,T}$, there exists a path-dependent symmetric threshold policy, which we denote $\mathcal{D}^o_{0,T}$, which does not underperform $\mathcal{P}_{0,T}$. By Lemma III.1, if there exists such a $\mathcal{D}^o_{0,T}$, then there exists an optimal path-independent symmetric threshold policy $\mathcal{S}^*_{0,T}$. Consequently, by Proposition III.1, there exists a Kalman-like estimator $\mathcal{H}$ that is jointly optimal to our problem alongside $\mathcal{S}^*_{0,T}$.

We show that $\exists\mathcal{D}^o_{0,T}$, $\forall\mathcal{P}_{0,T}$ by induction, where we start by proving the statement for the case $T=1$.

Induction Basis

Construct a policy $\mathcal{D}^o_{0,1}$ as

$$r^0_1 \triangleq \begin{cases}1, & \text{if } |\check{e}_1|>\tau_1,\\ 0, & \text{otherwise}\end{cases},$$

where τ_1 is a threshold that we will select appropriately. Consider a policy $\mathcal{P}_{0,1}$ that can be expressed as

$$r_1 \triangleq \begin{cases}1, & \text{if } \check{e}_1\in\mathbb{S},\\ 0, & \text{otherwise}\end{cases},$$

where $\mathbb{S}\subset\mathbb{R}$ is a measurable set. We define the function $h:\mathbb{R}\to\{0,1\}$ to be the indicator function of the set $\mathbb{S}$.

We start by noticing that for $\mathcal{P}_{0,1}$ and $\mathcal{D}^o_{0,1}$, it holds that

$$\gamma_{1|0}\equiv\gamma^o_{1|0}\equiv\mathcal{N}_{\sigma^2_w}, \tag{20}$$

where the superscript o denotes that the conditional PDF stands for the policy $\mathcal{D}^o_{0,1}$. This is due to the assumption that $r_0=1$, and given the time evolution of $\check{e}_k$ as in (12). In addition, from Proposition III.2 we have that the cost associated with the policy $\mathcal{P}_{0,1}$ is

$$J_{0,1}(\mathcal{P}_{0,1})=\mathbb{E}_{\gamma_{1|1}}\left[(\check{e}_k-\mathbb{E}_{\gamma_{1|1}}[\check{e}_2|\mathcal{I}^{\mathrm{e}}_k])^2\right]\varsigma_k+c(1-\varsigma_1).$$

From (19),

$$\gamma_{1|1}=\frac{\gamma_{1|0}h}{\varsigma_{1|0}}. \tag{21}$$

We then construct a desirable $\mathcal{D}^o_{0,1}$ by selecting τ_1 such that $\varsigma^o_1=\varsigma_1$. Similarly to finding (21), we have that

$$\gamma^o_{1|1}=\begin{cases}\frac{\gamma^o_{1|0}}{\varsigma^o_1}, & \text{if } |\check{e}_1|\leq\tau_1,\\ 0, & \text{otherwise,}\end{cases}$$

which is both neat and even, due to the Gaussianity of $\gamma^o_{1|0}$ as given in (20). Note that given our set up, we have that

$$\int_{\mathbb{R}}\gamma_{1|0}(e)h(e)\mathrm{d}e=\int_{-\tau}^{\tau}\gamma_{1|1}(e)\mathrm{d}e=\varsigma_1,$$

and therefore, by Lemma C.2,

$$\gamma^o_{1|1}\succ\frac{\gamma_{1|0}h}{\varsigma_1}=\gamma_{1|1}. \tag{22}$$

Consequently, from Lemma C.4, we have

$$\begin{aligned}&\mathbb{E}_{\gamma^o_{1|1}}\left[\left(\check{e}_1-\mathbb{E}_{\gamma^o_{1|1}}[\check{e}_1|\mathcal{I}^{\mathrm{e}}_1]\right)^2\right]\\ &\leq\mathbb{E}_{\gamma_{1|1}}\Bigg[\Big(\check{e}_1-\mathbb{E}_{\gamma^o_{1|1}}[\check{e}_1|\mathcal{I}^{\mathrm{e}}_1]\\ &\qquad-\left(\mathbb{E}_{\gamma_{1|1}}[\check{e}_1|\mathcal{I}^{\mathrm{e}}_1]-\mathbb{E}_{\gamma^o_{1|1}}[\check{e}_1|\mathcal{I}^{\mathrm{e}}_1]\right)\Big)^2\Bigg]\\ &=\mathbb{E}_{\gamma_{1|1}}\left[\left(\check{e}_1-\mathbb{E}_{\gamma_{1|1}}[\check{e}_1|\mathcal{I}^{\mathrm{e}}_1]\right)^2\right].\end{aligned} \tag{23}$$

Since the cost associated with the policy $\mathcal{D}^o_{0,1}$ is given by

$$J_{0,1}(\mathcal{D}^o_{0,1})=\mathbb{E}^o_{\gamma_{1|1}}\left[(\check{e}_k-\mathbb{E}^o_{\gamma_{1|1}}[\check{e}_2|\mathcal{I}^{\mathrm{e}}_1])^2\right]\varsigma_k+c(1-\varsigma_1),$$

we conclude from (23) that

$$J_{0,1}(\mathcal{P}_{0,1})\geq J_{0,1}(\mathcal{D}^o_{0,1}).$$

Therefore, $\mathcal{D}^o_{0,1}$ does not underperform $\mathcal{P}_{0,1}$.

Induction Step

Now given any $T>1$, assume the inductive hypothesis that $\mathcal{D}^o_{0,t}$ does not underperform $\mathcal{P}_{0,t}$ $\forall t<T$, which implies that there exist $\mathcal{S}^*_{1,T}$ through $\mathcal{S}^*_{T,T}$ that satisfy

$$J_{m,T}(\mathcal{S}^*_{m,T})=\min_{\mathcal{P}_{m,T}\in\mathbb{P}_{T-m}}J_{m,T}(\mathcal{P}_{m,T})=J^*_{m,T}, \quad m\leq T, \tag{24}$$

where $\mathcal{S}^*_{m,T}$ is of the symmetric threshold type $\mathbb{S}_{T-m}$ and the last equality follows from Definition III.6.

We define the following algorithm for $\mathcal{D}^o_{0,T}$:

Description of the Algorithm for $\mathcal{D}^o_{0,T}$.

- (Initial Step) Set $k=m$, and transmit the current state, i.e., $s_0=x_0$, or equivalently set $\check{e}_0=0$.

- (Step A) Increase the counter k by one. If $k > T$ holds then terminate, otherwise execute Step B.
- (Step B) If $|\check{e}_k| < \tau_k^o$ holds, then set $r_k = 0$, transmit the free symbol (i.e., $s_k = \mathfrak{E}$), and return to Step A. If $|\check{e}_k| \geq \tau_k^o$ holds, then execute $\mathcal{S}^*_{k,T}$, as defined in (24), where $\tau^o_{1:T}$ are appropriately chosen thresholds.

End of the description of the Algorithm for $\mathcal{D}^o_{0,T}$.

Starting from (22), whenever we have that

$$\gamma^o_{k|k} \succ \gamma_{k|k}$$

we use the time evolution of the PDF of $\check{e}_k$ as shown in (18), Lemmas C.3 and C.5 to get that

$$\gamma^o_{k+1|k} \succ \gamma_{k+1|k}.$$

In addition, whenever we have that

$$\gamma^o_{k+1|k} \succ \gamma_{k+1|k},$$

we proceed as before where we set τ_k^o such that $\varsigma_k^o = \varsigma_k$, and by using Lemma C.2, we conclude that

$$\gamma^o_{k+1|k+1} \succ \gamma_{k+1|k+1}.$$

By repeatedly applying this, we have that $\gamma^o_{k|k} \succ \gamma_{k|k}$ for all k in $\{1, \dots, T\}$. In addition, since $\gamma^0_{k|k}$ maintains zero-mean Gaussianity within the bounds $[-\tau, \tau]$, and is 0 otherwise, it is neat and even. Lemma C.4 then implies that

$$\begin{aligned}
&\mathbb{E}_{\gamma^o_{k|k}}\left[\left(\check{e}_k - \mathbb{E}_{\gamma^o_{k|k}}[\check{e}_k|\mathcal{I}^{\mathrm{e}}_k]\right)^2\right] \\
&\leq \mathbb{E}_{\gamma_{k|k}}\left[\left(\check{e}_k - \mathbb{E}_{\gamma^o_{k|k}}[\check{e}_k|\mathcal{I}^{\mathrm{e}}_k]\right.\right. \\
&\quad \left.\left. - \left(\mathbb{E}_{\gamma_{k|k}}[\check{e}_k|\mathcal{I}^{\mathrm{e}}_k] - \mathbb{E}_{\gamma^o_{k|k}}[\check{e}_k|\mathcal{I}^{\mathrm{e}}_k]\right)\right)^2\right] \\
&= \mathbb{E}_{\gamma_{k|k}}\left[\left(\check{e}_k - \mathbb{E}_{\gamma_{k|k}}[\check{e}_k|\mathcal{I}^{\mathrm{e}}_k]\right)^2\right]. \qquad (25)
\end{aligned}$$

By constructing the cost function from Proposition III.2 of any arbitrary policy and comparing it to that of $\mathcal{D}^o_{0,T}$, (25) gives

$$J_{0,T}(\mathcal{D}^o_{0,T}) \leq J_{0,T}(\mathcal{P}_{0,T}).$$

which proves the theorem. □

B. Noisy, Partial-State Feedback

Now address the general case with no further assumptions on (1) and (2), based on the work of [7]. A main assumption is that feedback is available from the estimator to the sensor. In addition, let the transmission be of the innovation $\mathbf{z}_k$ at time step k, i.e.,

$$\mathbf{s}_k = \mathbf{z}_k.$$

The goal of this paper is to derive the MMSE estimator for deterministic communication policies. The main results will be summarized in Theorem III.2.

By the definition of $\check{\mathbf{P}}_k$ and $\mathbf{R}$, there exists a unitary matrix $\mathbf{U}_k \in \mathbb{R}^{m\times m}$ such that

$$\mathbf{U}_k^\mathsf{T}(\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T} + \mathbf{R})\mathbf{U}_k = \mathbf{\Lambda}_k, \qquad (26)$$

where $\mathbf{\Lambda}_k = \mathrm{diag}(\lambda_k^1, \dots, \lambda_k^m) \in \mathbb{R}^{m\times m}$ and $\lambda_k^1, \dots, \lambda_k^m \in \mathbb{R}$ are the eigenvalues of $\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T} + \mathbf{R}$. Define $\mathbf{F}_k \in \mathbb{R}^{m\times m}$ as

$$\mathbf{F}_k \triangleq \mathbf{U}_k\mathbf{\Lambda}_k^{-1/2}. \qquad (27)$$

Lemma III.2. Given $\mathbf{F}_k$ defined as above, we have

$$(\mathbf{F}_k\mathbf{F}_k^\mathsf{T})^{-1} = \mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T} + \mathbf{R}.$$

Proof. We have that

$$\mathbf{\Lambda}_k^{\frac{1}{2}} = \mathbf{F}_k^{-1}\mathbf{U}_k,$$

therefore,

$$\mathbf{\Lambda}_k = \mathbf{U}_k^\mathsf{T}\mathbf{F}_k^{-1^\mathsf{T}}\mathbf{F}_k^{-1}\mathbf{U}_k = \mathbf{U}_k^\mathsf{T}(\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T} + \mathbf{R})\mathbf{U}_k,$$

which gives the equality $(\mathbf{F}_k\mathbf{F}_k^\mathsf{T})^{-1} = \mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T} + \mathbf{R}$. □

The matrix $\mathbf{F}_k$ is computed by the remote estimator and is sent back to the sensor along with $\mathbf{C}\check{\mathbf{x}}_k$ at each time. By defining $\boldsymbol{\epsilon_k}$ as

$$\boldsymbol{\epsilon}_k \triangleq \mathbf{F}_k^\mathsf{T}\mathbf{z}_k, \qquad (28)$$

which is known as the *Mahalanobis transformation*, we consider a deterministic event-based scheduler of the form

$$\gamma_k = \begin{cases} 0, & \text{if } \|\boldsymbol{\epsilon}_k\|_\infty \leq \delta, \\ 1, & \text{otherwise,} \end{cases} \qquad (29)$$

where $\delta \geq 0$ is a fixed threshold. Note that under such a scheduler, the estimator can infer that $\|\boldsymbol{\epsilon}_k\|_\infty \leq \delta$ when no signal is received, which helps reduce the estimation error at the remote estimator.

The MMSE estimate is uniquely specified as the conditional mean given all available information. We first provide an exact MMSE estimator corresponding to the event-based scheduler (29), and then derive a linear estimator from a Gaussianity assumption.

1) Exact MMSE Estimator: The *a priori* estimate $\check{\mathbf{x}}_k$ is

$$\check{\mathbf{x}}_k = \mathbb{E}[\mathbf{x}_k|\mathcal{I}^{\mathrm{e}}_{k-1}] = \int_{\mathbb{R}^n}\mathbf{x}f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathrm{d}\mathbf{x}. \qquad (30)$$

The corresponding estimation error covariance $\check{\mathbf{P}}_k$ is given by

$$\check{\mathbf{P}}_k = \int_{\mathbb{R}^n}(\mathbf{x} - \check{\mathbf{x}}_k)(\mathbf{x} - \check{\mathbf{x}}_k)^\mathsf{T}f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathrm{d}\mathbf{x}.$$

The *a posteriori* estimate $\hat{\mathbf{x}}_k$, which is the conditional mean of $\mathbf{x}_k$ given by $\mathcal{I}^{\mathrm{e}}_k$, is derived as follows. We have the following two cases:

1) $\underline{\gamma_k = 0}$: The sensor does not send $\mathbf{z}_k$ to the remote estimator, but the estimator is aware that $\|\boldsymbol{\epsilon}_k\|_\infty \leq \delta$. Consequently, $\hat{\mathbf{x}}_k$ is given by

$$\hat{\mathbf{x}}_k = \mathbb{E}[\mathbf{x}_k|\hat{\mathcal{I}}_k] = \int_{\mathbb{R}^n}\mathbf{x}f_{\mathbf{x}_k}(\mathbf{x}|\hat{\mathcal{I}}_k)\mathrm{d}\mathbf{x},$$

where $\hat{\mathcal{I}}_k \triangleq \mathcal{I}^{\mathrm{e}}_{k-1} \cup \{\lambda_k = 0\}$. The *a posteriori* error covariance $\mathbf{P}_k$ is given by

$$\mathbf{P}_k = \int_{\mathbb{R}^n} (\mathbf{x}-\hat{\mathbf{x}}_k)(\mathbf{x}-\hat{\mathbf{x}}_k)^\mathsf{T} f_{\mathbf{x}_k}(\mathbf{x}|\hat{\mathcal{I}}_k)\mathrm{d}\mathbf{x}.$$

Define the set $\Omega \subset \mathbb{R}^m$ as

$$\Omega \triangleq \{\boldsymbol{\epsilon}_k \in \mathbb{R}^m : \|\boldsymbol{\epsilon}_k\|_\infty \le \delta\},$$

then one can compute $f_{\mathbf{x}_k}(\mathbf{x}|\hat{\mathcal{I}}_k)$ using Bayes' rule as

$$\begin{aligned} f_{\mathbf{x}_k}(\mathbf{x}|\hat{\mathcal{I}}_k) &= f_{\mathbf{x}_k}\left(\mathbf{x}|\lambda_k = 0, \mathcal{I}^{\mathrm{e}}_{k-1}\right) \\ &= \frac{f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1}) p_{\lambda_k}(\lambda_k = 0|\mathbf{x}_k, \mathcal{I}^{\mathrm{e}}_{k-1})}{p_{\lambda_k}(\lambda_k = 0|\mathcal{I}^{\mathrm{e}}_{k-1})} \\ &= \frac{f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1}) \mathbb{P}\left(\|\boldsymbol{\epsilon}_k\|_\infty \le \delta \middle| \mathbf{x}_k, \mathcal{I}^{\mathrm{e}}_{k-1}\right)}{\mathbb{P}\left(\|\boldsymbol{\epsilon}_k\|_\infty \le \delta \middle| \mathcal{I}^{\mathrm{e}}_{k-1}\right)} \\ &= \frac{f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1}) \int_\Omega f_{\boldsymbol{\epsilon}_k}\left(\boldsymbol{\epsilon}_k \middle| \mathbf{x}_k, \mathcal{I}^{\mathrm{e}}_{k-1}\right)}{\int_\Omega f_{\boldsymbol{\epsilon}_k}\left(\boldsymbol{\epsilon}_k \middle| \mathcal{I}^{\mathrm{e}}_{k-1}\right)}. \end{aligned}$$

Given (2) and (6), we have that

$$\mathbf{z}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k - \mathbf{C}\check{\mathbf{x}}_k = \mathbf{C}\check{\mathbf{e}}_k + \mathbf{v}_k. \tag{31}$$

Consequently, given (28), we have that

$$f_{\boldsymbol{\epsilon}_k}\left(\boldsymbol{\epsilon}_k \middle| \mathbf{x}_k, \mathcal{I}^{\mathrm{e}}_{k-1}\right) \equiv \mathcal{N}(\mathbf{F}_k^\mathsf{T}\mathbf{C}(\mathbf{x}_k - \check{\mathbf{x}}_k), \mathbf{F}_k^\mathsf{T}\mathbf{R}\mathbf{F}_k),$$

which comes from the fact that $\boldsymbol{\epsilon}_k$ has an m-variable standard Gaussian distribution. In addition, using marginalization,

$$f_{\boldsymbol{\epsilon}_k}\left(\boldsymbol{\epsilon}_k \middle| \mathcal{I}^{\mathrm{e}}_{k-1}\right) = \int_{\mathbb{R}^n} f_{\boldsymbol{\epsilon}_k}\left(\boldsymbol{\epsilon}_k \middle| \mathbf{x}_k, \mathcal{I}^{\mathrm{e}}_{k-1}\right) f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathrm{d}\mathbf{x}.$$

2) $\underline{\gamma_k = 1}$: The sensor sends $\mathbf{z}_k$ to the remote estimator; therefore, $\mathcal{I}^{\mathrm{e}}_k$ becomes

$$\mathcal{I}^{\mathrm{e}}_k = \mathcal{I}^{\mathrm{e}}_{k-1} \cup \{\mathbf{s}_k = \boldsymbol{z}_k, \gamma_k = 1\}.$$

The remote estimator updates $\hat{\mathbf{x}}_k$ as per

$$\hat{\mathbf{x}}_k = \mathbb{E}[\mathbf{x}_k|\mathcal{I}_k] = \int_{\mathbb{R}^n} \mathbf{x} f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_k)\mathrm{d}\mathbf{x}.$$

The *a posteriori* error covariance $\mathbf{P}_k$ is given by

$$\mathbf{P}_k = \int_{\mathbb{R}^n} (\mathbf{x}-\hat{\mathbf{x}}_k)(\mathbf{x}-\hat{\mathbf{x}}_k)^\mathsf{T} f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_k)\mathrm{d}\mathbf{x}.$$

As before, using Bayes's rule, we can find that

$$f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_k) = \frac{f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1}) f_{\mathbf{z}_k}(\boldsymbol{z}|\mathcal{I}^{\mathrm{e}}_{k-1}, \mathbf{x}_k)}{f_{\mathbf{z}_k}(\boldsymbol{z}|\mathcal{I}^{\mathrm{e}}_{k-1})},$$

where

$$f_{\mathbf{z}_k}(\boldsymbol{z}|\mathcal{I}^{\mathrm{e}}_{k-1}, \mathbf{x}_k) \equiv \mathcal{N}(\mathbf{C}(\mathbf{x}_k - \check{\mathbf{x}}_k), \mathbf{R})$$

and

$$f_{\mathbf{z}_k}(\boldsymbol{z}|\mathcal{I}^{\mathrm{e}}_{k-1}) = \int_{\mathbb{R}^n} f_{\mathbf{z}_k}(\boldsymbol{z}|\mathcal{I}^{\mathrm{e}}_{k-1}, \mathbf{x}_k) f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathrm{d}\mathbf{x}.$$

Remark III.4. Note that the theory behind Kalman filter is developed for the case when $\gamma_k = 1$, $\forall k$, which justifies the assumption that $f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}_{k-1})$ is Gaussian given the dynamics. However, when $\gamma_k = 0$, this assumption is no longer valid, since now the filter has knowledge that the measurement was within some range in the previous time-steps, and this breaks the Gaussianity of $f_{\mathbf{x}_k}(\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1})$.

Remark III.5. Although the above two steps produce the MMSE estimate $\hat{\mathbf{x}}_k$ corresponding to the event-based scheduler (29), each updating step requires numerical integration. The amount of computation involved makes this estimator intractable in general, which motivates us to consider an approximate MMSE estimator.

2) Approximate MMSE Estimator: A commonly used approximation technique in nonlinear filtering is to assume that the conditional distribution of $\mathbf{x}_k$ given $\mathcal{I}_{k-1}$ is Gaussian, i.e.,

$$f_{\mathbf{x}_k}(\mathbf{x}_k|\mathcal{I}_{k-1}) \equiv \mathcal{N}(\check{\mathbf{x}}_k, \check{\mathbf{P}}_k). \tag{32}$$

This assumption reduces the estimation problem from the tracking of a general PDF, which is usually computationally intractable, to the tracking of its mean and covariance matrix. The approximation leads to a simple, linear form of the estimator, but the authors do not consider what was mentioned by Remark III.4, which might result in a significantly poor assumption.

Define the functions $\boldsymbol{h}, \boldsymbol{g}_\lambda, \tilde{\boldsymbol{g}}_\lambda : \mathbb{S}^n_+ \to \mathbb{S}^n_+$ as

$$\begin{aligned} \boldsymbol{h}(\mathbf{X}) &\triangleq \mathbf{A}\mathbf{X}\mathbf{A}^\mathsf{T} + \mathbf{Q}, \\ \tilde{\boldsymbol{g}}_\lambda(\mathbf{X}) &\triangleq \mathbf{X} - \lambda \mathbf{X}\mathbf{C}^\mathsf{T}[\mathbf{C}\mathbf{X}\mathbf{C}^\mathsf{T} + \mathbf{R}]^{-1}\mathbf{C}\mathbf{X}, \\ \boldsymbol{g}_\lambda(\mathbf{X}) &\triangleq \tilde{\boldsymbol{g}}_\lambda(\mathbf{X}) \circ \boldsymbol{h}(\mathbf{X}), \end{aligned}$$

Theorem III.2. Consider the remote state estimation with the event-based scheduler (29). Under the assumption (32), the MMSE estimator is given recursively as follows:

1) *Time update*:

$$\begin{aligned} \check{\mathbf{x}}_k &= \mathbf{A}\hat{\mathbf{x}}_{k-1}, \\ \check{\mathbf{P}}_k &= \boldsymbol{h}(\mathbf{P}_{k-1}). \end{aligned}$$

2) *Measurement update*:

$$\begin{aligned} \hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \gamma_k \mathbf{L}_k \mathbf{z}_k, \\ \check{\mathbf{P}}_k &= \gamma_k \tilde{\boldsymbol{g}}_1(\check{\mathbf{P}}_k) + (1-\gamma_k)\tilde{\boldsymbol{g}}_{\beta(\delta)}(\check{\mathbf{P}}_k), \end{aligned}$$

where

$$\beta(\delta) = \frac{2}{\sqrt{2\pi}}\delta e^{-\frac{\delta^2}{2}}[1 - 2Q(\delta)]^{-1}, \tag{33}$$

$Q(\cdot)$ is the standard Q-function defined by

$$Q(\delta) \triangleq \int_\delta^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\mathrm{d}x, \tag{34}$$

and $\mathbf{L}_k$ is the Kalman gain

$$\mathbf{L}_k = \check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}[\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T} + \mathbf{R}]^{-1}.$$

Before we present the proof, we state a few preliminary results. From the time update (30), the value of $\mathbf{z}_k$ as given in (31), and the assumption (32), we have that $\mathbf{z}_k$ is zero-mean Gaussian conditioned on $\mathcal{I}^{\mathrm{e}}_{k-1}$. Furthermore, assuming the estimation error is independent of the state given $\mathcal{I}^{\mathrm{e}}_{k-1}$, we have that $\mathbf{z}_k$ is jointly Gaussian with $\mathbf{x}_k$ conditioned on $\mathcal{I}^{\mathrm{e}}_{k-1}$.

From (31),

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{z}_k\mathbf{z}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}\right]\\
&= \mathbb{E}\left[(\mathbf{C}\check{\mathbf{e}}_\mathbf{k}+\mathbf{v}_\mathbf{k})(\check{\mathbf{e}}_k^\mathsf{T}\mathbf{C}^\mathsf{T}+\mathbf{v}_k^\mathsf{T})|\mathcal{I}^{\mathrm{e}}_{k-1}\right]\\
&= \mathbb{E}[\mathbf{C}\check{\mathbf{e}}_k\check{\mathbf{e}}_k^\mathsf{T}\mathbf{C}^\mathsf{T}+\mathbf{C}\check{\mathbf{e}}_k\mathbf{v}_k^\mathsf{T}+\mathbf{v}_k\check{\mathbf{e}}_k^\mathsf{T}\mathbf{C}^\mathsf{T}+\mathbf{v}_k\mathbf{v}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}]\\
&= \mathbf{C}\mathbb{E}[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}]\mathbf{C}^\mathsf{T}+\mathbf{C}\mathbb{E}[\check{\mathbf{e}}_k\mathbf{v}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}]\\
&\qquad + \mathbb{E}[\mathbf{v}_k\check{\mathbf{e}}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}]\mathbf{C}^\mathsf{T}+\underbrace{\mathbb{E}[\mathbf{v}_k\mathbf{v}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}]}_{=\mathbb{E}[\mathbf{v}_k\mathbf{v}_k^\mathsf{T}]}\\
&= \mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}+\mathbf{R},
\end{aligned} \tag{35}$$

and

$$\begin{aligned}
\mathbb{E}\left[\check{\mathbf{e}}_k\mathbf{z}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}\right] &= \mathbb{E}\left[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^\mathsf{T}\mathbf{C}^\mathsf{T}+\check{\mathbf{e}}_k\mathbf{v}_k|\mathcal{I}^{\mathrm{e}}_{k-1}\right]\\
&= \mathbb{E}\left[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}\right]\mathbf{C}^\mathsf{T}+\mathbb{E}\left[\check{\mathbf{e}}_k\mathbf{v}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}\right]\\
&= \check{\mathbf{P}}_k\mathbf{C}^\mathsf{T},
\end{aligned} \tag{36}$$

where we have used the fact that the measurement noise is assumed to be independent of the state and the measurement noise at all previous time steps; therefore, the measurement noise is independent of the estimation error.

Using the definition of $\boldsymbol{\epsilon}_k$ in (28), we have that

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}] &= \mathbf{F}_k^\mathsf{T}\mathbb{E}[\mathbf{z}_k\mathbf{z}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1}]\mathbf{F}_k\\
&\overset{\text{(a)}}{=} \mathbf{F}_k^\mathsf{T}(\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}+\mathbf{R})\mathbf{F}_k\\
&\overset{\text{(b)}}{=} \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}_k^\mathsf{T}(\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}+\mathbf{R})\mathbf{U}_k\boldsymbol{\Lambda}^{-\frac{1}{2}}\\
&\overset{\text{(c)}}{=} \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-\frac{1}{2}}\\
&= \mathbf{I}_m,
\end{aligned}$$

where (a) comes from (35), (b) from (27), (c) from (26), and $\mathbf{I}_m$ denotes an identity matrix size m. Therefore, given $\mathcal{I}^{\mathrm{e}}_{k-1}$, $\boldsymbol{\epsilon}_k$ is a zero-mean Gaussian multivariate random variable with unit variance.

Denote ϵ_k^i as the i^{th} element of $\boldsymbol{\epsilon}_k$, then as per the properties of $\boldsymbol{\epsilon}_k$ arising from the Mahalanobis transformation in (28), ϵ_k^i and ϵ_k^j are mutually independent if $i \neq j$. We then have the following result.

Lemma III.3.

$$\mathbf{F}_k^\mathsf{T}\mathbb{E}[\mathbf{z}_k\mathbf{z}_k^\mathsf{T}|\hat{\mathcal{I}}_k]\mathbf{F}_k = \mathbb{E}[\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\mathsf{T}|\hat{\mathcal{I}}_k] = [1-\beta(\delta)]\mathbf{I}_m.$$

Proof. The proof of the first equality is straightforward. Using Lemma D.1 and the mutual independence of the elements of $\boldsymbol{\epsilon}_k$, we have that

$$\begin{aligned}
\mathbb{E}\left[\left(\epsilon_k^i\right)^2|\hat{\mathcal{I}}_k\right] &= \mathbb{E}\left[\left(\epsilon_k^i\right)^2|\mathcal{I}^{\mathrm{e}}_{k-1},\|\boldsymbol{\epsilon}_k\|_\infty\le\delta\right]\\
&= \mathbb{E}\left[\left(\epsilon_k^i\right)^2|\mathcal{I}^{\mathrm{e}}_{k-1},|\epsilon_k^i|_\infty\le\delta\right]\\
&= 1-\beta(\delta),
\end{aligned}$$

and

$$\mathbb{E}[\epsilon_k^i\epsilon_k^j|\hat{\mathcal{I}}_k] = 0.$$

Therefore,

$$\mathbb{E}[\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\mathsf{T}|\hat{\mathcal{I}}_k] = [1-\beta(\delta)]\mathbf{I}_m.$$

□

In addition, the following lemma is used in deriving the main result.

Lemma III.4. The following equalities hold:

$$\mathbb{E}\left[\check{\mathbf{e}}_k\mathbf{z}_k^\mathsf{T}|\hat{\mathcal{I}}_k\right] = \mathbf{L}_k\mathbb{E}\left[\mathbf{z}_k\mathbf{z}_k^\mathsf{T}|\hat{\mathcal{I}}_k\right], \tag{37}$$

$$\mathbb{E}\left[(\check{\mathbf{e}}_k-\mathbf{L}_k\mathbf{z}_k)\mathbf{z}_k^\mathsf{T}|\hat{\mathcal{I}}_k\right] = 0, \tag{38}$$

$$\mathbb{E}\left[(\check{\mathbf{e}}_k-\mathbf{L}_k\mathbf{z}_k)(\check{\mathbf{e}}_k-\mathbf{L}_k\mathbf{z}_k)^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1},\mathbf{z}_k=\boldsymbol{z}\right] = \tilde{\boldsymbol{g}}(\check{\mathbf{P}}_k), \tag{39}$$

$$\mathbb{E}\left[(\check{\mathbf{e}}_k-\mathbf{L}_k\mathbf{z}_k)(\check{\mathbf{e}}_k-\mathbf{L}_k\mathbf{z}_k)^\mathsf{T}|\hat{\mathcal{I}}_k\right] = \tilde{\boldsymbol{g}}(\check{\mathbf{P}}_k), \tag{40}$$

where $\mathbf{L}_k = \check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}\left[\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}+\mathbf{R}\right]^{-1}$.

Proof. We have that $\mathbb{E}[\mathbf{x}_k|\mathcal{I}^{\mathrm{e}}_{k-1}] = \check{\mathbf{x}}_k$ and $\mathbb{E}[\mathbf{z}_k|\mathcal{I}^{\mathrm{e}}_{k-1}] = 0$. Using (35) and (36), and the results of Lemma D.2, we have that

$$\mathbb{E}[\mathbf{x}_k|\mathcal{I}^{\mathrm{e}}_{k-1},\mathbf{z}_k=\boldsymbol{z}] = \check{\mathbf{x}}_k+\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}[\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}+\mathbf{R}]^{-1}\boldsymbol{z}. \tag{41}$$

We define

$$p_\delta \triangleq \mathbb{P}[\|\boldsymbol{\epsilon}_k\|_\infty\le\delta|\mathcal{I}^{\mathrm{e}}_{k-1}].$$

Therefore,

$$\begin{aligned}
f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\hat{\mathcal{I}}_k) &= f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}^{\mathrm{e}}_{k-1},\gamma_k=0)\\
&= f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}^{\mathrm{e}}_{k-1},\|\boldsymbol{\epsilon}_k\|_\infty\le\delta)\\
&= \frac{f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathbb{P}[\|\boldsymbol{\epsilon}_k\|_\infty\le\delta|\boldsymbol{\epsilon},\mathcal{I}^{\mathrm{e}}_{k-1}]}{\mathbb{P}[\|\boldsymbol{\epsilon}_k\|_\infty\le\delta|\mathcal{I}^{\mathrm{e}}_{k-1}]}\\
&= \begin{cases}\frac{f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}^{\mathrm{e}}_{k-1})}{p_\delta}, & \text{if } \|\boldsymbol{\epsilon}\|_\infty\le\delta,\\ 0, & \text{otherwise.}\end{cases}
\end{aligned} \tag{42}$$

We can then prove (37) by showing that

$$\begin{aligned}
&\mathbb{E}[\check{\mathbf{e}}_k\mathbf{z}_k^\mathsf{T}|\hat{\mathcal{I}}_k]\\
&= \mathbb{E}[\check{\mathbf{e}}_k\mathbf{z}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1},\|\boldsymbol{\epsilon}\|\le\delta]\\
&= \mathbb{E}\left[\mathbb{E}[\check{\mathbf{e}}_k\mathbf{z}_k^\mathsf{T}|\mathcal{I}^{\mathrm{e}}_{k-1},\|\boldsymbol{\epsilon}\|\le\delta,\mathbf{z}_k=\boldsymbol{z}]|\hat{\mathcal{I}}_k\right]\\
&= \mathbb{E}\left[\mathbb{E}[\check{\mathbf{e}}_k|\mathcal{I}^{\mathrm{e}}_{k-1},\|\boldsymbol{\epsilon}\|\le\delta,\mathbf{z}_k=\boldsymbol{z}]\mathbf{z}_k^\mathsf{T}|\hat{\mathcal{I}}_k\right]\\
&\overset{\text{(a)}}{=} \mathbb{E}\left[\mathbb{E}[\check{\mathbf{e}}_k|\mathcal{I}^{\mathrm{e}}_{k-1},\mathbf{z}_k=\mathbf{F}_k^{-1^\mathsf{T}}\boldsymbol{\epsilon}]\boldsymbol{\epsilon}^\mathsf{T}\mathbf{F}_k^{-1}|\hat{\mathcal{I}}_k\right]\\
&= \int_\Omega \mathbb{E}[\check{\mathbf{e}}_k|\mathcal{I}^{\mathrm{e}}_{k-1},\mathbf{z}_k=\mathbf{F}_k^{-1^\mathsf{T}}\boldsymbol{\epsilon}]\boldsymbol{\epsilon}^\mathsf{T}\mathbf{F}_k^{-1}f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\hat{\mathcal{I}}_k)\mathrm{d}\boldsymbol{\epsilon}\\
&\overset{\text{(b)}}{=} \frac{1}{p_\delta}\int_\Omega \mathbb{E}[\check{\mathbf{e}}_k|\mathcal{I}^{\mathrm{e}}_{k-1},\mathbf{z}_k=\mathbf{F}_k^{-1^\mathsf{T}}\boldsymbol{\epsilon}]\boldsymbol{\epsilon}^\mathsf{T}\mathbf{F}_k^{-1}f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathrm{d}\boldsymbol{\epsilon}\\
&= \frac{1}{p_\delta}\int_\Omega \mathbb{E}[\mathbf{x}-\check{\mathbf{x}}_k|\mathcal{I}^{\mathrm{e}}_{k-1},\mathbf{z}_k=\mathbf{F}_k^{-1^\mathsf{T}}\boldsymbol{\epsilon}]\boldsymbol{\epsilon}^\mathsf{T}\mathbf{F}_k^{-1}f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathrm{d}\boldsymbol{\epsilon}\\
&= \frac{1}{p_\delta}\int_\Omega \left[\mathbb{E}[\mathbf{x}|\mathcal{I}^{\mathrm{e}}_{k-1},\mathbf{z}_k=\mathbf{F}_k^{-1^\mathsf{T}}\boldsymbol{\epsilon}]-\check{\mathbf{x}}_k\right]\\
&\qquad\qquad \times\boldsymbol{\epsilon}^\mathsf{T}\mathbf{F}_k^{-1}f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathrm{d}\boldsymbol{\epsilon}\\
&\overset{\text{(c)}}{=} \frac{1}{p_\delta}\int_\Omega \left[\check{\mathbf{x}}_k+\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}[\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\mathsf{T}+\mathbf{R}]^{-1}\mathbf{F}_k^{-1^\mathsf{T}}\boldsymbol{\epsilon}-\check{\mathbf{x}}_k\right]\\
&\qquad\qquad \times\boldsymbol{\epsilon}^\mathsf{T}\mathbf{F}_k^{-1}f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}^{\mathrm{e}}_{k-1})\mathrm{d}\boldsymbol{\epsilon}
\end{aligned}$$

$$
\begin{aligned}
&= \check{\mathbf{P}}_k\mathbf{C}^\top[\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\top + \mathbf{R}]^{-1}\mathbf{F}_k^{-1^\top}\frac{1}{p_\delta}\int_\Omega \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}_{k-1}^{\mathrm{e}})\mathrm{d}\boldsymbol{\epsilon}\mathbf{F}_k^{-1} \\
&= \mathbf{L}_k\mathbf{F}_k^{-1^\top}\mathbb{E}[\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\top|\hat{\mathcal{I}}_k]\mathbf{F}_k^{-1} \\
&\overset{(\mathrm{d})}{=} \mathbf{L}_k\mathbb{E}[\mathbf{z}_k\mathbf{z}_k^\top|\hat{\mathcal{I}}_k],
\end{aligned}
$$

where (a) is from (28), (b) is from (42), (c) is from (41), and (d) is from Lemma III.3. Note that (38) comes directly from (37).

To be able to find (39), we have from (4), (35), and (36) that

$$
\begin{aligned}
&\mathbb{E}\left[(\mathbf{x}_k - \mathbb{E}[\mathbf{x}_k|\mathcal{I}_{k-1}^{\mathrm{e}}])(\mathbf{x}_k - \mathbb{E}[\mathbf{x}_k|\mathcal{I}_{k-1^{\mathrm{e}}}])^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right] \\
&\qquad\qquad\qquad\qquad = \mathbb{E}[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^\top|\mathcal{I}_{k-1}^{\mathrm{e}}] = \check{\mathbf{P}}_k, \\
&\mathbb{E}\left[(\mathbf{x}_k - \mathbb{E}[\mathbf{x}_k|\mathcal{I}_{k-1}^{\mathrm{e}}])\mathbf{z}_k^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right] = \mathbb{E}[\check{\mathbf{e}}_k\mathbf{z}_k^\top|\mathcal{I}_{k-1}^{\mathrm{e}}] = \check{\mathbf{P}}_k\mathbf{C}^\top, \\
&\mathbb{E}\left[\mathbf{z}_k\mathbf{z}_k^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right] = \mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\top + \mathbf{R}.
\end{aligned}
$$

Therefore, using Lemma D.2, we have that

$$
\begin{aligned}
&\mathbb{E}[(\mathbf{x}_k - \mathbb{E}[\mathbf{x}_k|\mathcal{I}_{k-1}^{\mathrm{e}}, \mathbf{z}_k = \boldsymbol{z}_k]) \\
&\qquad\qquad \times (\mathbf{x}_k - \mathbb{E}[\mathbf{x}_k|\mathcal{I}_{k-1}^{\mathrm{e}}, \mathbf{z}_k = \boldsymbol{z}_k])^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}, \mathbf{z}_k = \boldsymbol{z}_k] \\
&= \mathbb{E}[\hat{\mathbf{e}}_k\hat{\mathbf{e}}_k^\top|\mathcal{I}_{k-1}^{\mathrm{e}}, \mathbf{z}_k = \boldsymbol{z}_k] \\
&= \check{\mathbf{P}}_k - \check{\mathbf{P}}_k\mathbf{C}^\top(\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^\top + \mathbf{R})^{-1}\mathbf{C}\check{\mathbf{P}}_k \\
&= \tilde{\boldsymbol{g}}(\check{\mathbf{P}}_{\mathbf{k}}),
\end{aligned}
$$

and from (41),

$$
\mathbf{x}_k - \mathbb{E}[\mathbf{x}_k|\mathcal{I}_{k-1}^{\mathrm{e}}, \mathbf{z}_k = \boldsymbol{z}_k] = \mathbf{x}_k - \check{\mathbf{x}}_k - \mathbf{L}_k\boldsymbol{z} = \check{\mathbf{e}}_k - \mathbf{L}_k\boldsymbol{z},
$$

which proves (39).

Using (39) and the same strategy as was used to prove (40), we have that

$$
\begin{aligned}
&\mathbb{E}\left[(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k)(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k)^\top\big|\hat{\mathcal{I}}_k\right] \\
&= \int_\Omega \mathbb{E}\left[(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k)(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k)^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}, \mathbf{z}_k = \mathbf{F}_k^{-1^\top}\boldsymbol{\epsilon}\right] \\
&\qquad\qquad\qquad\qquad\qquad \times \frac{f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}_{k-1}^{\mathrm{e}})}{p_\delta}\mathrm{d}\boldsymbol{\epsilon} \\
&= \tilde{\boldsymbol{g}}(\check{\mathbf{P}}_k),
\end{aligned} \tag{43}
$$

which proves (40). $\square$

Therefore, we now have all the preliminaries to prove Theorem III.2.

Proof. (Proof of Theorem III.2) The proof of the time update is simple. Given the process model (1),

$$
\begin{aligned}
\check{\mathbf{x}}_k &= \mathbb{E}[\mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}|\mathcal{I}_{k-1}^{\mathrm{e}}] \\
&= \mathbf{A}\mathbb{E}[\mathbf{x}_{k-1}|\mathcal{I}_{k-1}^{\mathrm{e}}] + \cancelto{0}{\mathbb{E}[\mathbf{w}_{k-1}|\mathcal{I}_{k-1}^{\mathrm{e}}]} \\
&= \mathbf{A}\hat{\mathbf{x}}_{k-1},
\end{aligned}
$$

and

$$
\begin{aligned}
\check{\mathbf{P}}_k &= \mathbb{E}\left[(\mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} - \check{\mathbf{x}}_k)\left(\mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} - \check{\mathbf{x}}_k\right)^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right] \\
&= \mathbb{E}\left[(\mathbf{A}\hat{\mathbf{e}}_{k-1} + \mathbf{w}_{k-1})\left(\mathbf{A}\hat{\mathbf{e}}_{k-1} + \mathbf{w}_{k-1}\right)^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right] \\
&= \mathbb{E}[\mathbf{A}\mathbf{e}_{\mathbf{k-1}}\hat{\mathbf{e}}_{k-1}^\top\mathbf{A}^\top + \mathbf{A}\hat{\mathbf{e}}_{k-1}\mathbf{w}_{k-1}^\top + \mathbf{w}_{k-1}\hat{\mathbf{e}}_{k-1}^\top\mathbf{A}^\top \\
&\qquad\qquad\qquad\qquad + \mathbf{w}_{k-1}\mathbf{w}_{k-1}^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}] \\
&= \mathbf{A}\mathbb{E}\left[\hat{\mathbf{e}}_{k-1}\hat{\mathbf{e}}_{k-1}^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right]\mathbf{A}^\top + \mathbf{A}\cancelto{0}{\mathbb{E}\left[\hat{\mathbf{e}}_{k-1}\mathbf{w}_{k-1}^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right]} \\
&\quad + \cancelto{0}{\mathbb{E}\left[\mathbf{w}_{k-1}\hat{\mathbf{e}}_{k-1}^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right]}\mathbf{A}^\top + \underbrace{\mathbb{E}\left[\mathbf{w}_{k-1}\mathbf{w}_{k-1}^\top\big|\mathcal{I}_{k-1}^{\mathrm{e}}\right]}_{=\mathbb{E}\left[\mathbf{w}_{k-1}\mathbf{w}_{k-1}^\top\right]} \\
&= \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^\top + \mathbf{Q} \\
&= \boldsymbol{h}(\mathbf{P}_{k-1}).
\end{aligned}
$$

Next, we verify the measurement update for the following two cases.

1) $\underline{\gamma_k = 1}$: From (41) and (43), we have

$$
\begin{aligned}
\hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \mathbf{L}_k\mathbf{z}_k, \\
\mathbf{P}_k &= \tilde{\mathbf{g}}(\check{\mathbf{P}}_k).
\end{aligned}
$$

2) $\underline{\gamma_k = 0}$: The sensor does not sent information regarding the measurement $\mathbf{x}_k$ at time k to the remote estimator, which computes $\hat{\mathbf{x}}_k$ as

$$
\begin{aligned}
\hat{\mathbf{x}}_k &= \mathbb{E}[\mathbf{x}_k|\hat{\mathcal{I}}_k] \\
&\overset{(\mathrm{a})}{=} \frac{1}{p_\delta}\int_\Omega \mathbb{E}\left[\mathbf{x}_k|\mathcal{I}_{k-1}^{\mathrm{e}}, \mathbf{z}_k = \mathbf{F}_k^{-1^\top}\boldsymbol{\epsilon}\right] f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}_{k-1}^{\mathrm{e}})\mathrm{d}\boldsymbol{\epsilon} \\
&\overset{(\mathrm{b})}{=} \frac{1}{p_\delta}\int_\Omega \left(\check{\mathbf{x}}_k + \mathbf{L}_k\mathbf{F}_k^{-1^\top}\boldsymbol{\epsilon}\right) f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}_{k-1}^{\mathrm{e}})\mathrm{d}\boldsymbol{\epsilon} \\
&= \frac{1}{p_\delta}\check{\mathbf{x}}_k\int_\Omega f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}_{k-1}^{\mathrm{e}})\mathrm{d}\boldsymbol{\epsilon} \\
&\qquad\qquad + \frac{\mathbf{L}_k\mathbf{F}_k^{-1^\top}}{p_\delta}\int_\Omega \boldsymbol{\epsilon} f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}_{k-1}^{\mathrm{e}})\mathrm{d}\boldsymbol{\epsilon},
\end{aligned}
$$

where (a) uses the same procedure as was used in the proof of (37), and (b) uses (41),

We have shown before that $\boldsymbol{\epsilon}_k$ is a zero-mean conditional Gaussian given $\mathcal{I}_{k-1}^{\mathrm{e}}$, and since by definition Ω is symmetric about 0, we have that

$$
\int_\Omega \boldsymbol{\epsilon} f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}_{k-1}^{\mathrm{e}})\mathrm{d}\boldsymbol{\epsilon} = 0.
$$

In addition, by definition,

$$
\int_\Omega f_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}|\mathcal{I}_{k-1}^{\mathrm{e}})\mathrm{d}\boldsymbol{\epsilon} = p_\delta.
$$

Therefore, the remote estimator's estimate simplifies to

$$
\hat{\mathbf{x}}_k = \check{\mathbf{x}}_k.
$$

The corresponding error covariance matrix $\mathbf{P}_k$ is given by

$$
\begin{aligned}
\mathbf{P}_k &= \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}}_k)(\mathbf{x} - \hat{\mathbf{x}}_k)^\top|\hat{\mathcal{I}}_k] \\
&= \mathbb{E}[(\mathbf{x} - \check{\mathbf{x}}_k)(\mathbf{x} - \check{\mathbf{x}}_k)^\top|\hat{\mathcal{I}}_k] \\
&= \mathbb{E}[((\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k) + \mathbf{L}_k\mathbf{z}_k)(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k) + \mathbf{L}_k\mathbf{z}_k)^\top|\hat{\mathcal{I}}_k] \\
&= \mathbb{E}[(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k)(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k)^\top|\hat{\mathcal{I}}_k] \\
&\qquad\qquad + \mathbb{E}[(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k)\mathbf{z}_k^\top|\hat{\mathcal{I}}_k]\mathbf{L}_k^\top \\
&\qquad\qquad + \mathbf{L}_k\mathbb{E}[\mathbf{z}_k(\check{\mathbf{e}}_k - \mathbf{L}_k\mathbf{z}_k)^\top|\hat{\mathcal{I}}_k]
\end{aligned}
$$

$$
\begin{aligned}
&+ \mathbf{L}_k \mathbb{E}[\mathbf{z}_k \mathbf{z}_k^\mathsf{T} | \hat{\mathcal{I}}_k] \mathbf{L}_k^\mathsf{T} \\
&\overset{\text{(a)}}{=} \tilde{\boldsymbol{g}}(\check{\mathbf{P}}_k) + \mathbf{L}_k \mathbf{F}_k^{-1^\mathsf{T}} \mathbb{E}[\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^\mathsf{T} |_k] \mathbf{F}_k^{-} \mathbf{L}_k^\mathsf{T} \\
&\overset{\text{(b)}}{=} \tilde{\boldsymbol{g}}(\check{\mathbf{P}}_k) + (1 - \beta(\delta)) \mathbf{L}_k (\mathbf{F}_k \mathbf{F}_k^\mathsf{T})^{-1} \mathbf{L}_k^\mathsf{T} \\
&\overset{\text{(c)}}{=} \tilde{\boldsymbol{g}}(\check{\mathbf{P}}_k) + (1 - \beta(\delta)) \mathbf{L}_k (\mathbf{C} \check{\mathbf{P}}_k \mathbf{C}^\mathsf{T} + \mathbf{R}) \mathbf{L}_k^\mathsf{T} \\
&= \tilde{\mathbf{g}}_{\beta(\delta)}(\check{\mathbf{P}}_k),
\end{aligned}
$$

where (a) is from (28), (38), and (40), (b) is from Lemma III.3, and (c) is from Lemma III.2, and this concludes the proof.

□

Remark III.6. $\mathbf{P}_k$ is a function of $\gamma_{0:k}$ and $\beta(\delta)$, both of which depend on δ. By properly tuning δ, we can achieve a desired trade-off between the sensor communication rate γ and the estimation quality in terms of $\mathbf{P}_k$. The authors do not address how to choose δ; however, it is kept general enough to implement any communication policy as long as it is of the threshold type, and is deterministic.

Lemma III.5. Let $\delta \geq 0$. Then

$$\mathbb{P}(\|\boldsymbol{\epsilon}_k\|_\infty \leq \delta | \mathcal{I}_{k-1}^{\mathrm{e}}) = [1 - 2Q(\delta)]^m.$$

Proof. Note that $\|\boldsymbol{\epsilon}_k\|_\infty = \max\{|\epsilon_k^1|, \dots |\epsilon_k^m|\} \leq \delta$ if and only if $\left||\epsilon_k^i| \leq \delta\right|, \forall 1 \leq i \leq m$. Since they are all i.i.d., with a zero-mean standard Gaussian,

$$
\begin{aligned}
\mathbb{P}(\|\boldsymbol{\epsilon}_k\|_\infty \leq \delta | \mathcal{I}_{k-1}^{\mathrm{e}}) &= \prod_{i=1}^{m} \mathbb{P}(\left||\epsilon_k^i|\right| \leq \delta | \mathcal{I}_{k-1}^{\mathrm{e}}) \\
&= \prod_{i=1}^{m} \int_{-\delta}^{\delta} f_{\epsilon_k^i}(\epsilon) \mathrm{d}\epsilon \\
&= (1 - 2Q(\delta))^m.
\end{aligned}
$$

□

Proposition III.3. Consider the remote state estimation problem with the event-based sensor schedule (29). Under the assumption (32), the expected per-step sensor-to-estimator communication cost $\bar{J}_{\mathrm{c}}(\Theta)$ is given by

$$\bar{J}_{\mathrm{c}}(\Theta) = 1 - [1 - 2Q(\delta)]^m.$$

Proof. From Lemma III.5, we have that γ is independent from $\mathcal{I}_{k-1}^{\mathrm{e}}$, and

$$
\begin{aligned}
\mathbb{P}(\gamma_k = 0) &= [1 - 2Q(\delta)]^m, \\
\mathbb{P}(\gamma_k = 1) &= 1 - [1 - 2Q(\delta)]^m.
\end{aligned}
$$

Therefore,

$$\bar{J}_{\mathrm{c}}(\Theta) = \mathbb{E}^{\Theta}\left[\gamma_k \middle| \mathcal{I}_{k-1}^{\mathrm{e}}\right] = \mathbb{E}^{\Theta}\left[\gamma_k\right] = 1 - [1 - 2Q(\delta)]^m.$$

□

IV. Stochastic Communication Policy

Despite the optimality of deterministic communication policies, the estimators prove to be complex and intractable. A Gaussian assumption breaks the optimality of the estimator, as explained in Remark III.4. Despite allowing reasonable estimates for small update sequences, the Gaussianity assumption results in poor performance when the communication scheme results in long periods of time without correction steps, as will be discussed in Section V. This motivates the study of stochastic communication policies.

The strategy behind developing a stochastic communication policy is to introduce external uncertainty in the system such that the derivation of the optimal estimator is feasible with a closed-form solution. To do so, the sensor generates independent and identically distributed (i.i.d.) random variables ζ_k at every time step k. These random variables are uniformly distributed over $[0, 1]$ and the sensor then compares ζ_k with a function $c_k(\mathbf{y}_k, \mathcal{I}_{k-1}^{\mathrm{s}})$, where in this case the function maps to the closed set $[0, 1]$. Therefore,

$$\gamma_k = \begin{cases} 0, & \text{if } \zeta_k \leq c_k(\mathbf{y}_k, \mathcal{I}_{k-1}^{\mathrm{s}}), \\ 1, & \text{otherwise.} \end{cases}$$

Remark IV.1. Note that one can interpret $c_k(\mathbf{y}_k, \mathcal{I}_{k-1}^{\mathrm{s}})$ as the probability the sensor will not transmit, and only by a careful selection of $c_k(\mathbf{y}_k, \mathcal{I}_{k-1}^{\mathrm{s}})$ does the existence of a tractable MMSE estimator occur.

We now briefly consider the work in [8], which addresses systems of the form (1) and (2). The authors assume that the sensor is of the primitive type, i.e., it does not have the processing capability to generate Kalman-like filter estimates, which is the case for many real-life problems.

The proofs will not be discussed due to the lack of space. In addition, the authors study two cases of the estimation problem; an open-loop system where only the raw measurement can be accessed by the sensor, and another closed-loop system where the sensor receives the estimate data of the estimator. For a fair comparison with the approximate MMSE estimator developed in Section III-B, which assumes a closed-loop system, we exclusively address the closed-loop system.

Given that the sensor receives feedback from the estimator, the transmitted signal $\mathbf{s}_k$ can be assumed to be the innovation, i.e.,

$$\mathbf{s}_k = \mathbf{z}_k,$$

where $\mathbf{z}_k$ is defined as in (6). As a result, we set up the stochastic communication to be

$$\gamma_k = \begin{cases} 0, & \text{if } \zeta_k \leq \nu(\mathbf{z}_k), \\ 1, & \text{otherwise,} \end{cases} \tag{44}$$

where

$$\nu(\mathbf{z}_k) \triangleq \exp\left(-\frac{1}{2}\mathbf{z}^\mathsf{T}\mathbf{Z}\mathbf{z}\right), \tag{45}$$

with $\mathbf{Z} \in \mathbb{S}_{++}^n$ being a "tuning parameter" specified by the user, which can be thought of as an additional degree of freedom to allocate resources between the communication rate and the estimation performance. In [8], details on how to optimize this parameter to satisfy different criteria are discussed. Note that, given (44), the larger $\mathbf{z}_k$ is, the larger is the probability that the sensor transmits to the estimator.

In addition, as with the deterministic policy case, when the estimator does not receive a signal from the sensor, it can exploit the information that $\mathbf{z}_k$ is more likely to be small to update the state estimate.

Theorem IV.1. Consider the remote state estimation problem with the closed-loop event-triggered schedule (44). Then $\mathbf{x}_k$ conditioned on $\mathcal{I}^{\mathrm{e}}_{k-1}$ is Gaussian distributed with mean $\check{x}_k$ and corresponding covariance $\check{\mathbf{P}}$, and $\mathbf{x}_k$ conditioned on $\mathcal{I}^{\mathrm{e}}_{k}$ is Gaussian distributed with mean $\hat{\mathbf{x}}_k$ and covariance $\mathbf{P}_k$, where $\check{\mathbf{x}}_k$, $\check{\mathbf{P}}_k$, $\hat{\mathbf{x}}_k$, and $\mathbf{P}_k$ satisfy the following recursive equations:

- *Time Update:*

$$\begin{aligned}\check{\mathbf{x}}_k &= \mathbf{A}\hat{\mathbf{x}}_{k-1},\\ \check{\mathbf{P}}_k &= \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^{\mathsf{T}} + \mathbf{Q}.\end{aligned}$$

- *Measurement Update:*

$$\begin{aligned}\hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \gamma_k \mathbf{K}_k \mathbf{z}_k\\ \mathbf{P}_k &= \check{\mathbf{P}}_k - \mathbf{K}_k\mathbf{C}\check{\mathbf{P}}_k,\end{aligned}$$

  where

$$\mathbf{K}_k = \check{\mathbf{P}}_k\mathbf{C}^{\mathsf{T}}\left[\mathbf{C}\check{\mathbf{P}}_k\mathbf{C}^{\mathsf{T}} + \mathbf{R} + (1-\gamma_k)\mathbf{Z}^{-1}\right]^{-1}$$

  with initial condition

$$\check{\mathbf{x}}_0 = 0, \quad \check{\mathbf{P}}_0 = \mathbf{\Sigma}_0.$$

Proof. The interested reader is referred to [8]. □

The results of Theorem IV.1 show that the propagation and update step is the same as the standard Kalman filter when $\gamma_k = 1$. In addition, when $\gamma_k = 0$, The *a posteriori* error covariance is inflated using an enlarged measurement noise covariance $\mathbf{R} + \mathbf{Z}^{-1}$.

An extension of this work involves a modified policy for system with smart sensors [9], which can communicate an estimate,

$$\mathbf{s}_k = \mathbb{E}[\mathbf{x}_k | \mathcal{I}^{\mathrm{s}}_{k-1}].$$

The authors show that with a stochastic event-based scheduling policy with a modified $c_k(\mathbf{y}_k, \mathcal{I}^{\mathrm{s}}_{k-1})$ results in closed-form expression of the MMSE estimate, which is the local Kalman filter at the estimator when a measurement is received and an inflated covariance when no measurement is received, due to the performance loss in the absence of communication. They then pose the problem of identifying optimal paramaters for the sensor scheduler as a dynamic programming (DP) problem, efficiently allocating communication resources over a finite time-horizon.

Other extensions to the case of multisensors [11] and nonlinear [12] set ups exist in literature, where in [12], the authors use an unscented kalman filter (UKF) for the estimation problem.

V. Simulations

In this section, we address the performance of the addressed remote estimation strategies. The work by [2] addressed in Section III-A proves that for any admissible communication policy, there exist a set of thresholds such that a symmetric threshold communication policy with these thresholds does not underperform the original, admissible communication policy. However, the authors do not provide a comprehensive strategy for achieving the ultimately optimal thresholds. In addition, the results are limited to perfect state-feedback systems; therefore, the work of [2] is excluded from the subsequent analysis.

The goal is then to analyze the performance of the remote estimation strategies developed in [7] and [8], which were discussed in Sections III-B and IV, respectively. To do so, we simulate a system of the form (1) and (2), with

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0.9 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix},$$
$$\mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 2 \end{bmatrix}.$$

In addition, for every simulation, we evaluate a cost function of the form

$$J_{0,T} = \log\left(\sum_{k=1}^{T} \mathrm{Tr}(\mathbf{P}_k)\right),$$

where $T = 10000$ seconds. We define

$$\Gamma = \sum_{k=1}^{T} 1 - \gamma_k,$$

i.e., Γ is the number of times the scheduler decides not to transmit.

We then present two baselines to compare the results to.

1) Standard Kalman filter: This is the case where the sensor communicates to the estimator at all times, i.e., $\gamma_k = 1$, $\forall k$, therefore resulting in state propagation and update as per the standard Kalman filter, given in [13].

2) No communication: This is the case where the sensor never transmits to the estimator, i.e., $\gamma_k = 0, \forall k$, therefore resulting in no state correction and only state propagation.

For the deterministic communication policy, Fig. 1 shows that the approximate MMSE estimator developed in Section III-B can result in superior estimation performance for low values of the parameter δ. However, δ increasing results in less transmission from the sensor to the transmitter, as shown in Fig. 2. Therefore, the poor performance at higher δ goes back to Remark III.4. As the update step is not triggered as often, the Gaussian approximation becomes less accurate for approximating the iteratively increasing nonlinearities, and thus resulting in a poorer estimate.

Meanwhile, Fig. 1 shows that the stochastic communication policy introduced in Section IV, despite the introduced uncertainty associated with the uniformly distributed random variable ζ_k, can always produce decent estimates comparable to the standard Kalman filter. Note that, for the purposes of these simulations, we vary the user-defined parameter $\mathbf{Z}$ in (45) by setting it to be

$$\mathbf{Z} = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix}.$$

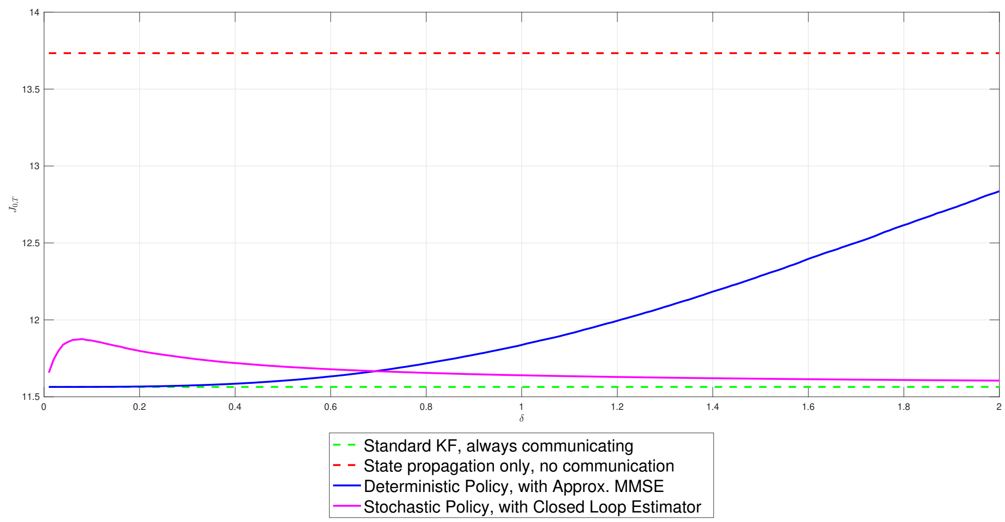

Fig. 1: A plot of the cost function of the two different communication policies, plotted against the value of the user-defined parameters. For reference, two additional experiments were run with a standard KF for a constantly communicating policy and another where the sensor never transmits to the estimator.

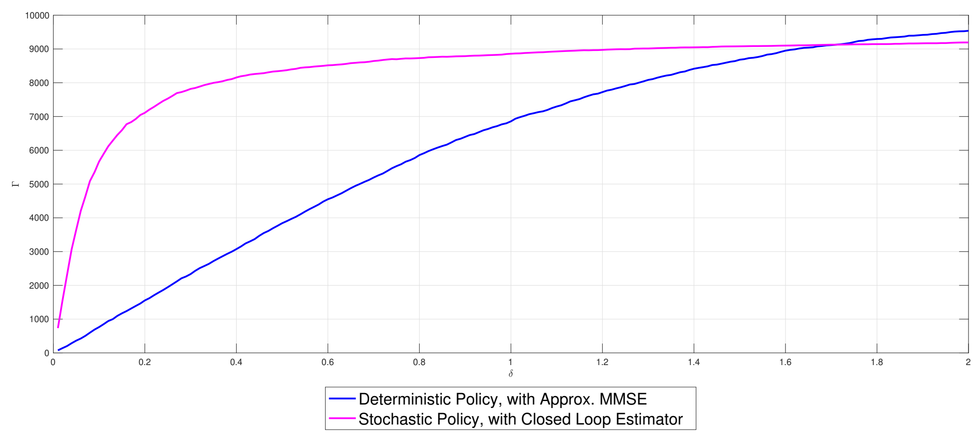

Fig. 2: A plot of Γ of the two different communication policies, plotted against the value of the user-defined parameters.

Note that for this particular reason, the cost value computed for the first estimator at a specific value of δ is not comparable to the cost value computed for the other estimator at that specific value of δ.

Given the stochastic communication policy, the *a priori* and *a posteriori* distributions of the state are conditionally Gaussian, and therefore, no information loss is encountered. Therefore, as shown by Figs. 1 and 2, the stochastic communication policy can handle long periods of time without a correction steps. By comparing the performance of the two estimators when the number of transmissions is roughly the same as seen in Fig, 2, the performance of the stochastic policy estimator is clearly superior.

Therefore, this motivates the following important conclusion. For systems with low communication cost, where δ will be tuned such that the scheduler transmits frequently to the estimator, a deterministic policy might be favourable as the approximation will be sufficiently accurate. However, in costly communication media, a stochastic policy will be advantageous as the lack of communication deems the Gaussian approximation required by the deterministic policy estimator weak.

VI. Conclusion

We have provided an overview of some of the keymark papers in remote estimation. We constrained our study exclusively to linear systems, with process noise. We have then posed the remote estimation as a problem of finding the joint optimal scheduler and estimator to minimize a cost function encompassing both an estimation error and a communication penalty.

We first showed that the work of [2] results in an optimal communication policy that is of the symmetric threshold type and the optimal estimator is Kalman-like. However, this search was restricted to systems of perfect state feedback. By introducing noisy measurements, and a feedback loop from the sensor to the estimator, general deteerministic policies were addressed to try and develop the MMSE estimator as per the work of [7]. The MMSE estimator was extracted, and a tractable approximate one was derived. We then addressed stochastic communication policies that introduce uncertainty through sampling from a uniform distribution and comparing it to a measure computed from the sensor innovation. We then stated a theorem that the MMSE estimator is tractable without the need for approximations, as given by [8].

Lastly, we compared the performance of the approximate MMSE estimator to that of the estimator for stochastic communication policies. We arrived at the conclusion that the deterministic policy estimator is superior because it can incorporate optimal communication policies without introducing additional uncertainties, and it can provide performance very close to that of the standard Kalman filter. However, this is the case when the number of communications is significant as it overcomes the inaccuracies introduced by the Gaussian assumption. When the estimator is expected to go through multiple iterations without receiving a measurement, then the penalty introduced by introducing uncertainty is overcome by the advantages of not assuming Gaussianity, and a stochastic communication policy then proves the better option.

It is clear that there is still room for further exploration though. To the best of my knowledge, there exists no deterministic policy with a tractable estimator that minimizes the total cost of the system (without costly approximations); therefore, there exists no comprehensive solution yet to the problem addressed in this paper. In addition, most of the literature focuses on simple linear systems, with the exception of the recent paper [12] that uses a UKF. The problem of standard nonlinear estimation has been extensively studied, but that has not yet yielded satisfactory results in the world of remote estimation.

Appendix A
Quasi-Convex Lemma

Lemma A.1. Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable, bounded, even and quasi convex-function. Let w be a random variable with an even and quasi-concave probability density function. Define $\bar{h} : \mathbb{R} \to \mathbb{R}$, such that $\bar{h} \triangleq \mathbb{E}[h(x + w)]$, then $\bar{h}$ is a bounded, even, and quasi-convex function. If the function h is also continuous then $\bar{h}$ is also continuous.

Proof. The interested reader is referred to [2]. □

Appendix B
Probability Theory

Lemma B.1. Let $x_1, \dots, x_N$ be independent random variables with marginal PDFs $f_{x_i}(x)$, $i = 1, \dots, N$, respectively. The PDF of $y = x_1 + \ldots + x_N$ is

$$f_y(y) = f_{x_1}(y) * \ldots * f_{x_N}(y),$$

where $*$ denotes convolution.

Proof. Using the definition of a characteristic function of a PDF, we have

$$\begin{aligned}\Phi_y(\omega) &= \mathbb{E}[e^{j\omega y}] \\ &= \mathbb{E}[e^{j(\omega x_1 + \ldots + \omega x_N)}] \\ &= \mathbb{E}[e^{j\omega x_1} \ldots e^{j\omega x_N}] \\ &= \Phi_{x_1}(\omega) \ldots \Phi_{x_N}(\omega).\end{aligned}$$

Applying the convolution theorem for Fourier transforms, we get

$$f_y(y) = f_{x_1}(y) * \ldots * f_{x_N}(y).$$

□

Appendix C
Results from Majorization Theory

The following Lemmas are stated without their proofs. The interested reader is referred to [2] and the references therein.

Lemma C.1. If f and h are neat and even probability density functions, then $f * h$ is also neat and even.

Let $\mathbb{A}$ be a given Borel measurable subset of $\mathbb{R}$, we denote its Lebesgue measure by $\mathcal{L}(\mathbb{A})$.

Definition C.1. (Symmetric Rearrangement) If the Lebesgue measure of $\mathbb{A}$ is finite, then the symmetric rearrangement of $\mathbb{A}$, denoted by $\mathbb{A}^\sigma$, is a symmetric closed interval centered around the origin with Lebesgue measure $\mathcal{L}(\mathbb{A})$

$$\mathbb{A}^\sigma = \left\{ x \in \mathbb{R} : |x| \leq \frac{\mathcal{L}(\mathbb{A})}{2} \right\}.$$

Definition C.2. Let $f; \mathbb{R} \to \mathbb{R}$ be a given non-negative function. We define f^σ, the symmetric non-decreasing rearrangement of f as

$$f^\sigma(x) \triangleq \int_0^\infty \mathcal{R}_{\{z \in \mathbb{R} : f(z) > p\}^\sigma}(x) \mathrm{d}\rho,$$

where $\mathcal{R}_{\{z \in \mathbb{R} : f(z) > p\}^\sigma} : \mathbb{R} \to \{0, 1\}$ is the following indicator function:

$$\mathcal{R}_{\{z \in \mathbb{R} : f(z) > p\}^\sigma}(x) \triangleq \begin{cases} 1, & \text{if } x \in \{z \in \mathbb{R} : f(z) > \rho\}^\sigma, \\ 0, & \text{otherwise}, \quad x \in \mathbb{R}. \end{cases}$$

Definition C.3. If g and g are two probability density functions on $\mathbb{R}$, then we say that f majorizes g, which we denote $f \succ g$, provided that the following holds

$$\int_{|x| \leq \rho} g^\sigma(x) \mathrm{d}x \leq \int_{|x| \leq \rho} f^\sigma(x) \mathrm{d}x, \qquad \forall \rho \geq 0.$$

Definition C.4. Given a probability density function $f : \mathbb{R} \to \mathbb{R}$ and a Borel set $\mathbb{K}$, such that $\int_{\mathbb{K}} f(x) \mathrm{d}x > 0$, we define the restriction of f to $\mathbb{K}$ as

$$f_{\mathbb{K}}(x) \triangleq \begin{cases} \frac{f(x)}{\int_{\mathbb{K}} f(x) \mathrm{d}x}, & \text{if } x \in \mathbb{K}, \\ 0, & \text{otherwise} . \end{cases}$$

Lemma C.2. Let $f, g : \mathbb{R} \to \mathbb{R}$ be two probability density functions, such that f is neat and even and $f \succ g$. Let κ be a real number in the interval $\kappa \in (0,1)$. and let $\mathbb{A} = [-\tau, \tau]$ be the symmetric closed interval such that $\int_{-\tau}^{\tau} f(x)\mathrm{d}x = 1-\kappa$. For any function $h : \mathbb{R} \to [0,1]$ satisfying $\int_{\mathbb{R}} g(x)h(x)\mathrm{d}x = 1-\kappa$, the following holds:

$$f_{\mathbb{A}} \succ \frac{g \cdot h}{1-\kappa}.$$

Lemma C.3. Left f and g be two probability density functions on $\mathbb{R}$, with f symmetric non-increasing and $f \succ g$. For a symmetric non-increasing probability density function h, the following holds:

$$f * h \succ g * h.$$

Lemma C.4. Let f be a neat and even probability density function on the real line. Let g be a probability density function on the real satisfying $f \succ g$. Then the following holds:

$$\int_{\mathbb{R}} x^2 f(x)\mathrm{d}x \leq \int_{\mathbb{R}} (x-y)^2 g(x)\mathrm{d}x, \quad y \in \mathbb{R}.$$

Lemma C.5. Let $f_x, f_y : \mathbb{R} \to \mathbb{R}$ represent the probability density functions of random variables x and y, such that $f_x \succ f_y$. For any non-zero constant a,

$$f_{ax} \succ f_{ay}.$$

Appendix D

Lemma D.1. Let $x \in \mathbb{R}$ be a Gaussian scalar random variable (r.v.) with zero mean and variance $\mathbb{E}[x^2] = \sigma^2$. Denoting $\Delta = \delta\sigma$, then $\mathbb{E}[x^2|\,|x| \leq \Delta] = \sigma^2(1-\beta(\delta))$, where $\beta(\delta)$ is defined as in (33).

Proof. We have that

$$f_x(x|\,|x| \leq \Delta) = \frac{f_x(x)}{\int_{-\Delta}^{\Delta} f_x(t)dt},$$

therefore,

$$\begin{aligned}\mathbb{E}[x^2|\,|x| \leq \Delta] &= \frac{1}{\int_{-\Delta}^{\Delta} f_x(t)\mathrm{d}t} \int_{-\Delta}^{\Delta} \frac{t^2}{\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} \mathrm{d}t \\ &= \frac{1}{\int_{-\Delta}^{\Delta} f_x(\sigma y)\mathrm{d}(\sigma y)} \int_{-\Delta}^{\Delta} \frac{(\sigma y)^2}{\sqrt{2\pi}} e^{-\frac{(\sigma y)^2}{2\sigma^2}} \mathrm{d}(\sigma y) \\ &= \frac{\sigma^2}{\int_{-\delta}^{\delta} f_x(y)\mathrm{d}y} \int_{-\delta}^{\delta} \frac{y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \mathrm{d}y,\end{aligned}$$

where we have used the transformation $t = \sigma y$. Using the Q-function defined in (34), we have that

$$\begin{aligned}\int_{-\delta}^{\delta} f_x(y)\mathrm{d}y &= 1 - 2\int_{\delta}^{\infty} f_x(y)\mathrm{d}y \\ &= 1 - 2Q(\delta).\end{aligned}$$

Therefore, we have that

$$\mathbb{E}[x^2|\,|x| \leq \Delta] = \frac{\sigma^2}{1-2Q(\delta)} \int_{-\delta}^{\delta} \frac{y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \mathrm{d}y.$$

Using integration by parts, we can find that

$$\int_{-\delta}^{\delta} \frac{y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \mathrm{d}y = 1 - 2Q(\delta) - \frac{2}{\sqrt{2\pi}} \delta e^{-\frac{\delta^2}{2}},$$

leading to Lemma D.1. □

Lemma D.2. (From [14]) Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ be jointly Gaussian with mean and variance

$$\mathbf{m} = \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}.$$

Then $\mathbf{x}$ is conditionally Gaussian given $\mathbf{y} = \boldsymbol{y}$ with

$$f_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y}) \equiv \mathcal{N}(\bar{\boldsymbol{x}} + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \bar{\boldsymbol{y}}), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}).$$

References

[1] Y. Xu and J. P. Hespanha, "Optimal communication logics in networked control systems," *Proceedings of the IEEE Conference on Decision and Control*, vol. 4, pp. 3527–3532, 2004.

[2] G. M. Lipsa and N. C. Martins, "Costs for First-Order LTI Systems," *IEEE transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, 2013.

[3] D. Ward and N. C. Martins, "Optimal Remote Estimation Over Use-Dependent Packet-Drop Channels," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 127–132, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.ifacol.2016.10.384

[4] J. Chakravorty and A. Mahajan, "Remote-state estimation with packet drop," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 7–12, 2016.

[5] S. Trimpe and R. D. Andrea, "Event-Based State Estimation With Variance-Based Triggering," *IEEE transactions on Automatic Control*, vol. 59, no. 12, pp. 3266–3281, 2014.

[6] J. Chakravorty and A. Mahajan, "Remote estimation over a packet-drop channel with Markovian state," *IEEE Transactions on Automatic Control*, pp. 1–1, 2019.

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

[8] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2661–2675, 2015.

[9] J. Wu, X. Ren, D. Han, D. Shi, and L. Shi, "Finite-horizon Gaussianity-preserving event-based sensor scheduling in Kalman filter applications," *Automatica*, vol. 72, pp. 100–107, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.automatica.2016.05.013

[10] S. Dey, A. Chiuso, and L. Schenato, "Remote estimation with noisy measurements subject to packet loss and quantization noise," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 204–217, 2014.

[11] S. Weerakkody, Y. Mo, B. Sinopoli, D. Han, and L. Shi, "Multi-Sensor Scheduling for State Estimation with Event-Based, Stochastic Triggers," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2695–2701, 2016.

[12] L. Li, D. Yu, Y. Xia, and H. Yang, "Remote Nonlinear State Estimation with Stochastic Event-Triggered Sensor Schedule," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 734–745, 2019.

[13] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[14] B. Anderson and J. Moore, *Optimal Filtering*. Prentice-Hall, 1979.