Event-Triggered State Estimation: An Iterative Algorithm and Optimality Properties

Adam Molin, *Member, IEEE,* and Sandra Hirche, *Senior Member, IEEE*

Abstract

This paper investigates the optimal design of event-triggered estimation for linear systems. The synthesis approach is posed as a team decision problem where the decision makers are given by the event-trigger and the estimator. The event-trigger decides upon its available measurements whether the estimator shall obtain the current state information by transmitting it through a resource constrained channel. The objective is to find the optimal trade-off between the mean square estimation error and the expected number of transmissions over a finite horizon. After deriving basic characteristics of the optimal solution, we propose an iterative algorithm that alternates between optimizing one decision maker while fixing the other and vice versa. By analyzing the dynamical behavior of the iterative method, it is shown that the algorithm converges to a symmetric threshold policy for first-order systems if the statistics of the uncertainties are even and unimodal. In the case of bimodal distributions, we show numerically that the iterative method may find asymmetric threshold policies that outperform symmetric rules.

Index Terms

Event-triggered state estimation, Networked control systems, Distributed decision making, Team optimality

A. Molin is with the Department of Automatic Control, School of Electrical Engineering, KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden; `adammol@kth.se`.

S. Hirche is with the Institute for Information-oriented Control, Technical University of Munich, D-80290 München, Germany; `hirche@tum.de`

I. INTRODUCTION

In contrast to the periodically sampled systems, in which measurements are taken within equidistant time-intervals, an event-triggered estimator receives measurement updates in an asynchronous fashion. The event-trigger is a preprocessing unit situated at the sensor which decides upon its available information, whether to update a remote estimator with new measurements. There is a lot of evidence in the literature that event-triggered sampling for estimation and control is a promising alternative to periodic sampling when the sensor transmissions are costly [1]–[7]. Along with the benefits of event-triggered sampling for estimation, several challenges emerge in the analysis and design that are not present in periodically sampled systems. One of the main issues corresponds to the implicit information available at the estimator in the absence of an event [8], [9]. This additional information commonly leads to a close coupling of the estimator and the event-trigger. In order to overcome this complication, the majority of results assumes symmetric threshold policies. The restriction to symmetry commonly implies that the estimator will not depend on the event-triggering rule [1]–[3]. In contrast to these works, the focus of this paper is to study exactly this dependence between event-trigger and estimator by allowing for asymmetric triggering policies.

Our approach is posed as a team decision problem composed of the event-trigger and the estimator. The optimal decision makers aim at minimizing a finite horizon cost criterion composed of the mean square estimation error and the expected number of transmissions. First, basic properties of the optimal event-trigger and estimator are derived in which we make use of the nested information pattern of the team decision problem. We then propose an iterative algorithm that alternates between minimizing the cost criterion over one decision maker while fixing the other and vice versa. Stationary solutions of this algorithm can be identified as person-by-person optimal solutions. By studying the convergence behavior of the proposed iterative algorithm, we are able to attain a more detailed characterization of the optimal solution under a slight restriction of the set of admissible policies. It is shown for first-order systems that the optimal event-trigger is given by a symmetric threshold policy if the distributions of the uncertainties are unimodal and symmetric. Furthermore, we demonstrate numerically that our algorithm yields asymmetric threshold policies for symmetric but bimodal noise densities. For certain systems, it turns out that these asymmetric solutions decrease the cost considerably when comparing with

the optimal symmetric solution. This observation suggests that unimodality of the distributions is essential for the optimality of symmetric threshold policies.

A. Related work

The optimality of symmetric threshold functions for the remote state estimation problem with costly communication has also been proven in [10]. The authors in [10] make use of majorization theory and the Riesz rearrangement inequality [11] in order to arrive at this result. The proof follows a similar guideline used for a related problem that studies the joint optimization of paging and registration policies in mobile networks [12]. The analysis of the asymptotic behavior of our iterative algorithm constitutes an alternative way to prove that symmetric event-triggering laws are optimal for first-order systems. In contrast to [10], our proposed method enables the analysis of multi-modal noise distributions and is capable to find solutions that outperform symmetric policies in this case.

Iterative methods for the solution of team decision problems, in which one policy is optimized while the others are fixed, has also been applied for the study of optimal solutions of the Witsenhausen's counter-example [13] and for the joint design of source-channel-relay mappings [14]. The work in [15] proposes an iterative encoder-decoder design algorithm for event-triggered feedback control over a bandwidth-limited channel. The authors in [16] focus on the existence of solutions of our proposed problem for higher-order systems by relating it to the Lloyds algorithm originally used to compute Centroidal Voronoi Tessellations [17].

Another closely related work [18] studies distributed estimation for two random variables over the collision channel. In contrast to our work, the estimator must distinguish between the absence of an event and a collision during transmission. The authors show that threshold policies are optimal for the case of a single snapshot estimation. Moreover, the results indicate that even in the case of observing random variables with unimodal and symmetric distributions, the optimal event-trigger thresholds may be asymmetric. The work in [19] considers a variation of our problem setup for first-order Markov sources. In [20], the authors consider the joint design of estimation and scheduling for an energy-harvesting sensor by employing results from majorization theory. Both works [19] and [20] highlight the importance of symmetry and unimodality.

A preliminary version of this work has been published in [21]. The present note differs from the work in [21] as follows. It presents the iterative method and establishes basic properties of

the optimal event-triggered estimator for higher-order systems. This is in contrast to [21], which merely considers first-order systems. Moreover, this note introduces a new crucial assumption necessary for a detailed study of optimal event-triggered estimation. Finally, it provides a full proof of Theorem 1 that has been omitted in [21].

B. Organization of the Paper

Sec. II introduces the design problem of event-triggered estimation under communication constraints. A basic characterization of the optimal solution is discussed in Sec. III. Based on these properties, an iterative algorithm is developed in Sec. IV whose convergence properties are analyzed in Sec. V.

C. Notation

The expected value is denoted by $\mathsf{E}^{\mathsf{f}}[\cdot]$ and the conditional expectation is denoted by $\mathsf{E}^{\mathsf{f}}[\cdot|\cdot]$, where the underlying probability measure P^{f} is parameterized by the policy f. The variable X^k denotes the sequence $[x_0, \dots, x_k]$ and X_k^l denotes the sequence $[x_k, \dots, x_l]$. The indicator function is denoted by $\mathbb{1}_A(x)$ taking a value of 1 if $x \in A$ and 0 otherwise. The complement of a set A is denoted by A^{c}. The maximum norm of a vector $x \in \mathbb{R}^n$ is denoted by $|x|_\infty$. The convolution of two real-valued function f and g is denoted by $\mathsf{f} * g$.

II. MMSE ESTIMATION UNDER COMMUNICATION CONSTRAINTS

Consider the following linear process $\mathcal{P}$ driven by noise w_k

$$x_{k+1} = Ax_k + w_k, \tag{1}$$

where x_k takes values in $\mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$. The system noise w_k takes values in $\mathbb{R}^n$ and is an i.i.d. random variable described by the probability density function ϕ_w, which is zero-mean and has a covariance matrix C_w. The initial state, x_0, is statistically independent of w_k and is described by probability density function ϕ_{x_0}, which has a finite mean $\mathsf{E}[x_0]$ and a covariance matrix C_{x_0}. The initial state x_0 and the noise process $\{w_k\}$ are also referred to as primitive random variables in the following. System parameters and statistics are known to both the event-trigger and estimator.

The system model is illustrated in Fig. 1. The process $\mathcal{P}$ outputs the state x_k. The event-trigger $\mathcal{E}$ decides upon its available information whether or not to transmit the current state to the remote state estimator $\mathcal{S}$. We define the output of the event-trigger as

$$\delta_k = \begin{cases} 1, & \text{update } x_k \text{ sent,} \\ 0, & \text{otherwise.} \end{cases}$$

The communication channel between the process $\mathcal{P}$ and the state estimator $\mathcal{S}$ can be viewed as a δ_k-controlled erasure channel whose outputs are described by

$$z_k = \begin{cases} x_k, & \delta_k = 1, \\ \varnothing, & \delta_k = 0, \end{cases} \tag{2}$$

where $\varnothing$ is the erasure symbol. As it will be useful for subsequent analysis, we define the last update time τ_k as

$$\tau_k = \max\{\kappa | \delta_\kappa = 1, \ \kappa < k\} \tag{3}$$

with $\tau_k = -1$, if no transmissions have occurred prior to k. The variable τ_k evolves by the δ_k-controlled difference equation

$$\tau_{k+1} = \begin{cases} k, & \delta_k = 1, \\ \tau_k, & \delta_k = 0, \end{cases} \qquad \tau_0 = -1. \tag{4}$$

Admissible event-triggers are given by causal mappings

$$\delta_k = \mathrm{f}_k(X^k), \quad k = 0, \ldots, N-1.$$

The estimator $\mathcal{S}$ outputs the state estimate $\hat{x}_k$ and is given by measurable mappings g_k defined by

$$\hat{x}_k = g_k(Z^k), \quad k = 0, \ldots, N-1.$$

The design objective is to jointly synthesize the event-trigger $\mathrm{f} = [\mathrm{f}_0, \ldots, \mathrm{f}_{N-1}]$ and the estimator $g = [g_0, \ldots, g_{N-1}]$ that minimize cost J defined by

$$J = \mathsf{E}^{\mathrm{f},g} \left[\sum_{k=0}^{N-1} \|x_k - \hat{x}_k\|^2 + \lambda \delta_k \right]. \tag{5}$$

The per-stage cost of J is composed of the squared estimation error $\|x_k - \hat{x}_k\|^2$ and a communication penalty $\lambda \delta_k$. The weight $\lambda > 0$ determines the amount of penalizing transmissions

over the communication channel emerging either from bandwidth limitations or from energy restrictions in the sensor node.
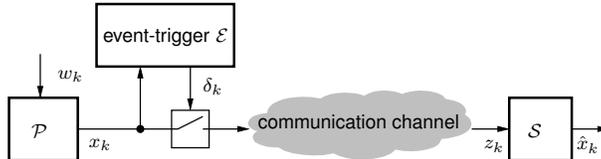

Fig. 1: System model of the networked estimation system with plant $\mathcal{P}$, event-trigger $\mathcal{E}$, state estimator $\mathcal{S}$ and communication channel.

III. BASIC PROPERTIES

In this section, we are concerned with finding basic properties of optimal solutions minimizing (5) that will facilitate the description of optimal event-triggered estimators. We begin with a characterization of the optimal estimator given an arbitrary event-trigger.

Proposition 1: For any event-trigger f, the optimal state estimator g^* is given by the MMSE estimator

$$\hat{x}_k = g_k^*(Z^k) = \mathsf{E}^{\mathrm{f}}[x_k|Z^k], \quad k = 0, \dots, N-1.$$

Proof: Fix an arbitrary event-trigger f. The sequence $\{\delta_k\}$ is then a function of primitive random variables. Hence, the communication penalty term $\mathsf{E}^{\mathrm{f}}\left[\sum_{k=0}^{N-1} \lambda \delta_k\right]$ is constant and can be omitted from the optimization. In the remaining estimation problem the mean square error $\mathsf{E}^{\mathrm{f}}\left[\sum_{k=0}^{N-1} \|x_k - \hat{x}_k\|^2\right]$ is to be minimized. The optimal solution for this problem is given by the MMSE estimator $\mathsf{E}^{\mathrm{f}}[x_k|Z^k]$, [22]. This completes the proof. ∎

In the following, we introduce a time-variant translatory change of coordinates of the state space that will enable us to focus on the main issues involved in the joint optimization of the event-trigger and the estimator. As the coordinate transformation at each time k can be computed at the event-trigger and the estimator, the optimization problem remains unchanged. Let us define the linear predictor $\hat{x}_k^{\mathrm{LP}}$ by the following recursion

$$\hat{x}_k^{\mathrm{LP}} = \begin{cases} x_k, & \delta_k = 1, \\ A\hat{x}_{k-1}^{\mathrm{LP}}, & \delta_k = 0, \end{cases} \tag{6}$$

for $k \in \{1, \ldots, N-1\}$ and $\hat{x}_0^{\mathrm{LP}} = \mathsf{E}[x_0]$. The linear predictor is the MMSE estimator, when having no information about the choice of the event-trigger f and assuming that transmission instances are statistically independent of the state evolution. Clearly, this holds when transmission instances are selected in advance.

Let us rewrite the optimization problem by defining the one-step ahead estimation error of the linear predictor as

$$e_k = x_k - A\hat{x}_{k-1}^{\mathrm{LP}}, \quad k = 1, \ldots, N-1 \tag{7}$$

and $e_0 = w_{-1}$, where we define $w_{-1} = x_0 - \mathsf{E}[x_0]$. The variable e_k defines our new state to be estimated and follows the recursion

$$e_{k+1} = h_k(e_k, \delta_k, w_k) = (1 - \delta_k)Ae_k + w_k. \tag{8}$$

Further, we define $\hat{e}_k$ to be the MMSE estimate $\mathsf{E}[e_k|\tilde{Z}^k]$, where $\tilde{z}_k$ is defined accordingly as

$$\tilde{z}_k = \begin{cases} e_k, & \delta_k = 1, \\ \varnothing, & \delta_k = 0. \end{cases} \tag{9}$$

It is straightforward to see that the estimation error $e_k - \hat{e}_k$ and $x_k - \hat{x}_k$ are identical random variables for a fixed event-trigger f, as e_k corresponds to a translatory coordinate transformation of x_k shifted by $-A\hat{x}_{k-1}^{\mathrm{LP}}$ which is known since the sequence δ^{k-1} is measurable with respect to Z^k. Therefore, our initial optimization problem with cost function J can be rewritten as

$$\inf_{\mathrm{f}} \mathsf{E}^{\mathrm{f}} \left[\sum_{k=0}^{N-1} \|e_k - \hat{e}_k\|^2 + \lambda \delta_k \right]. \tag{10}$$

In the following, we assume that the event-trigger f is given by a mapping from E^k to $\{0, 1\}$. Since there always exists a bijection from X^k to E^k given the variables $\delta_0, \ldots, \delta_{k-1}$, this change of variables does not put any restrictions on the further analysis keeping in mind that any policy expressed in E^k can also be written as a function in X^k. Let $\mathcal{F}$ denote the set of admissible policies over the horizon N.

In order to find further structural properties of the optimal solution, we need to introduce an assumption on our admissible event-triggering policies. Suppose $\delta_m = 1$ at time $m \in \{-1, \ldots, N-2\}$ and apply a policy f until time m. Then, the remaining optimization problem

is to find $f^m = [f_{m+1}, \ldots, f_{N-1}]$, that minimizes

$$J_m = \mathsf{E}^f \left[\sum_{\ell=m+1}^{N-1} \|e_\ell - \hat{e}_\ell\|^2 + \lambda \delta_\ell | \delta_m = 1 \right] \tag{11}$$

where e_ℓ evolves according to (8) with initial condition $e_{m+1} = w_m$. The above optimization problem has the same form as (10) for each m with a shifted initial condition and a shorter horizon. This suggests that optimal policies need not depend on data prior to time $\tau_k + 1$ because of the reset transition of the state variable e_m at $m = \tau_k$, i.e. it seems to be reasonable to restrict our policies to only depend on $E_{\tau_k+1}^k$. Though a further in-depth analysis for the verification of this assertion is needed as the expected value to be computed in (11) still depends on the complete policy f, this study exceeds the scope of this note. Instead, we provide the following restriction on the set of admissible policies, which our subsequent results will be based on.

Definition 1: The set $\mathcal{F}^{\text{CON}}$ defines the family of admissible event-triggering policies over the horizon N constrained to the form

$$\delta_k = f_k(E_{\tau_k+1}^k), \quad 0 \le k \le N - 1. \tag{12}$$

Based on the restricted set of policies introduced in Definition 1, the next proposition gives us insights into the structure of $\hat{e}_k$.

Proposition 2: Let the event-trigger $f \in \mathcal{F}^{\text{CON}}$ be fixed. Then, the MMSE estimate of e_k is given by

$$\hat{e}_k = \begin{cases} e_k, & \delta_k = 1, \\ \alpha_k(\tau_k), & \delta_k = 0, \end{cases} \tag{13}$$

where τ_k is defined by (3) and $\alpha_k(\tau_k)$ is defined by

$$\alpha_k(\tau_k) = \mathsf{E}^f \left[\sum_{l=\tau_k}^{k-1} A^{k-l-1} w_l | \delta_{\tau_k+1} = 0, \ldots, \delta_k = 0 \right]. \tag{14}$$

Proof: Clearly, we have $\hat{e}_k = e_k$ for $\delta_k = 1$, as $e_k \in \tilde{Z}^k$ if $\delta_k = 1$. For $\delta_k = 0$, τ_k is a sufficient statistics for $\hat{e}_k$ due to the state equation in (8) and $f \in \mathcal{F}^{\text{CON}}$. The mapping α_k is determined by applying recursively (8) with $e_{\tau_k+1} = w_{\tau_k}$. ∎

The function $\alpha = [\alpha_0, \ldots, \alpha_{N-1}]$ in Proposition 2 can be interpreted as a bias term to improve the state estimate by incorporating the negative information $\delta_{\tau_k+1} = \cdots = \delta_k = 0$ at time k.

Because of Proposition 2, we can rewrite optimization problem (10) as

$$\inf_{\mathrm{f}\in\mathcal{F}^{\mathrm{CON}}} \mathsf{E}^{\mathrm{f}}\left[\sum_{k=0}^{N-1}(1-\delta_k)\|e_k-\alpha_k(\tau_k)\|^2+\lambda\delta_k\right]. \tag{15}$$

It can be observed that the running cost reduces to λ and is therefore independent of the current α_k in the case $\delta_k=1$.

IV. AN ITERATIVE ALGORITHM

What prevents a further study of the optimization problem (15) is the fact that the estimation bias $\alpha_k(\tau_k)$ depends on the particular policy $\mathrm{f}\in\mathcal{F}^{\mathrm{CON}}$ chosen up to time k. Therefore, methods like dynamic programming are not directly applicable. In order to overcome this burden, we broaden (15) by considering the variable α_k as a new decision variable being a function of τ_k. Then, the optimization problem is given by

$$\inf_{\mathrm{f}\in\mathcal{F}^{\mathrm{CON}},\alpha} J \tag{16}$$

with

$$J(\mathrm{f},\alpha)=\mathsf{E}^{\mathrm{f}}\left[\sum_{k=0}^{N-1}(1-\delta_k)\|e_k-\alpha_k(\tau_k)\|^2+\lambda\delta_k\right]. \tag{17}$$

The optimization problem (16) enlarges the set of possible solutions compared to optimization problem (15), because it omits the constraint for α given by (14). On the other hand, because of Proposition 2, it follows by contradiction that the optimal solution of (16) satisfies (14). Because of these observations, we can conclude that any optimal solution of problem (16) will yield the same optimal cost as (15). By considering optimization problem (16), we are able to specify the structure of the optimal event-trigger given by the following proposition.
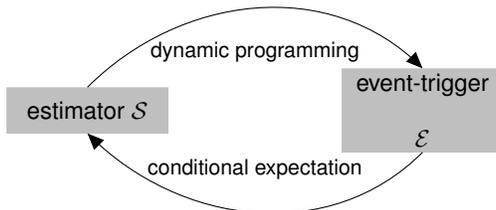

Fig. 2: Iterative scheme to calculate event-trigger $\mathcal{E}$ and estimator $\mathcal{S}$.

Proposition 3: Let the function α be fixed and the event-triggering law f be in $\mathcal{F}^{\mathrm{CON}}$. Then, the variables e_k and τ_k are a sufficient statistics for the optimal event-trigger f_k, $0 \le k \le N-1$.

Proof: The evolution of the pair (e_k, τ_k) can be regarded as a δ_k-controlled Markov process defined by (4) and (8). The running cost of J at time k is a function of the pair (e_k, τ_k), input δ_k and noise w_k. By [22], this problem can be solved by dynamic programming with (e_k, τ_k) being the state, which is a sufficient statistics of the optimal solution f_k. This completes the proof. ∎

Proposition 3 implies that the optimal event-trigger in $\mathcal{F}^{\mathrm{CON}}$ is a function of e_k and τ_k. It can be observed that for a fixed event-trigger f, the optimal map α can be calculated by (14). On the other hand, for any fixed map α, the optimal event-trigger f can be calculated by dynamic programming. We therefore define the running cost as

$$c_k^{\alpha_k}(e_k, \tau_k, \delta_k) = (1 - \delta_k)\|e_k - \alpha_k(\tau_k)\|^2 + \lambda \delta_k,$$

and the Bellman operator as

$$\mathcal{T}_k^{\alpha_k} J_{k+1}(\cdot) = \min_{\delta_k \in \{0,1\}} c_k^{\alpha_k}(\cdot, \delta_k) + \mathsf{E}\left[J_{k+1}(e_{k+1}, \tau_{k+1})|\cdot, \delta_k\right].$$

The value function J_k being a function of the augmented state (e_k, τ_k) is determined by recursive application of the Bellman equation given by

$$J_k = \mathcal{T}_k^{\alpha_k} J_{k+1}$$

with $J_N \equiv 0$, where the argument in the minimization yields the optimal event-trigger f and we have

$$J(f, \alpha) = \mathsf{E}^f[J_0(e_0, -1)].$$

This observation motivates us to propose the following iterative procedure sketched in Fig. 2, which alternates between optimizing f while fixing policy α and vice versa. Algorithm 1 describes the iterative procedure. With slight abuse of notation, we declared τ_k as a second subscript instead of an argument of α_k.

As the cost J decreases or is at least kept constant in each step of the iteration, the sequence $[(f^0, \alpha^0), (f^1, \alpha^1), \ldots]$ produces a non-increasing succession of costs J.

Algorithm 1 Iterative procedure to calculate (f, α)

Require: $\alpha^0_{k,\tau_k} \in \mathbb{R}, \quad k = 0, \ldots, N-1, \tau_k = -1, \ldots, k-1$

  1: $i \leftarrow 0$

  2: **repeat**

  3:     $k = N, \ J_N \equiv 0$

  4:     **repeat**

  5:        $k \leftarrow k - 1$

  6:        $J_k \leftarrow \mathcal{T}_k^{\alpha_k^i} J_{k+1}$

  7:        $f_k^i(e_k, \tau_k) \in \arg\min_{\delta_k \in \{0,1\}} c_k^{\alpha_k^i}(e_k, \tau_k, \delta_k) + \mathsf{E}\left[J_{k+1}(e_{k+1}, \tau_{k+1})|e_k, \tau_k, \delta_k\right]$

  8:     **until** $k = 0$

  9:     $\alpha^{i+1}_{k,\tau_k} \leftarrow \mathsf{E}^{f^i}\left[\sum_{l=\tau_k}^{k-1} A^{k-l-1} w_l | \delta^k_{\tau_k+1} = 0\right]$

10:     $i \leftarrow i + 1$

11: **until** convergence

V. CONVERGENCE PROPERTIES

In the following, we are interested in the convergence properties of the proposed iterative algorithm for scalar systems. We will therefore restrict our analysis to linear first-order processes $\mathcal{P}$ defined as

$$x_{k+1} = ax_k + w_k, \tag{18}$$

where $a \in \mathbb{R}\backslash\{0\}$. The system noise w_k takes values in $\mathbb{R}$ and is an i.i.d. random variable described by the probability density function ϕ_w, which is zero-mean and has finite variance. The initial state, x_0 is statistically independent of w_k and is described by density function ϕ_{x_0}, which has a finite mean $\mathsf{E}[x_0]$ and a finite variance. As in the previous sections, we will study the transformed system with state variable e_k defined in (7). Additionally, it is assumed that the density functions are symmetric around their means, i.e.,

$$\phi_w(w) = \phi_w(-w),$$

$$\phi_{e_0}(e) = \phi_{e_0}(-e)$$

for all $w, e \in \mathbb{R}$. Rather than regarding α as a function of k and τ_k, we will interpret α as a vector in $\mathbb{R}^{\frac{1}{2}N(N+1)}$ by reindexing its entries appropriately. A central notion for the optimality of two-person team problems is given by person-by-person optimality, which is defined as follows with regard to (16).

Definition 2 (Person-by-person optimality): A solution (α^*, f^*) of (16) is called *person-by-person optimal*, if

$$J(\alpha^*, \mathrm{f}^*) \leq J(\alpha, \mathrm{f}^*),$$

$$J(\alpha^*, \mathrm{f}^*) \leq J(\alpha^*, \mathrm{f})$$

for all $\alpha \in \mathbb{R}^{\frac{1}{2}N(N+1)}$ and all admissible policies f.

The above definition means to say that the cost of a person-by-person optimal solution can not be decreased by either changing the estimation bias α or by changing the event-trigger f while fixing the other. Person-by-person optimality is a necessary condition for optimality, since it would be otherwise possible to improve the solution by the iterative algorithm defined by Algorithm 1. Hence, every fixpoint (f^*, α^*) in the iterative algorithm is a person-by-person optimal solution of optimization problem (16).

The following proposition shows that $\alpha^* = 0$ and its resulting optimal event-trigger f^* is a person-by-person optimal solution.

Proposition 4: Let e_0 and $\{w_k\}$ have symmetric distributions. Then $\alpha^* = 0$ is a fixpoint of the Algorithm 1. The policy of the event-trigger f^* that corresponds to α^* is an even mapping of e_k and is independent of τ_k for every $k = 0, \ldots, N-1$.

Proof: Let α^0 be 0 for all k and all τ in the initialization of Algorithm 1. The cost function J reduces then to

$$J(\mathrm{f}, \alpha^0) = \mathsf{E}^{\mathrm{f}} \left[\sum_{k=0}^{N-1} (1-\delta_k)|e_k|^2 + \lambda\delta_k \right]$$

where e_k evolves by the recursion (8). Therefore, the resulting optimal f_k^0 and the cost-to-go function J_k is only a function of e_k for all $k = 0, \ldots, N-1$. We first show that the application of the Bellman operator $\mathcal{T}_k^0$ preserves symmetry of J_{k+1} for any k. Given an even value function J_{k+1}, the conditional expectation $\mathsf{E}[J_{k+1}(e_{k+1})|\cdot, \delta_k]$ preserves symmetry for both $\delta_k = 1$ as the expected value is constant. Due to the symmetry of ϕ_{w/e_0} and the system dynamics (8),

we have for $\delta_k = 1$

$$\mathsf{E}\left[J_{k+1}(ae_k + w_k)|e_k\right] = \mathsf{E}\left[J_{k+1}(-ae_k - w_k)|e_k\right] =$$

$$= \mathsf{E}\left[J_{k+1}(-ae_k + w_k)|e_k\right] = \mathsf{E}\left[J_{k+1}(ae_k + w_k)| - e_k\right]$$

Adding the cost $c_k^0(\cdot, \delta_k)$ also preserves symmetry, because the sum of two even functions is again even. Taking the pointwise minimum of two even functions yields an even function. Therefore, an even function remains even after application of the Bellman operator. As $J_N \equiv 0$ is an even function, it follows by induction that every value function J_k is even for $k \in \{0, \ldots, N-1\}$. This implies that f_k^0 resulting in the first iteration step from Algorithm 1 is an even mapping of e_k, if $\alpha^0 = 0$.

Next, we calculate α^1 assuming f_k^0 being an even function of e_k for $k \in \{0, \ldots, N-1\}$. Let $\phi_{e_k|\tau}$ be the density function of the conditional probability distribution of e_k given τ_k and $\delta_k = 0$. Then, due to (14) and Proposition 2, $\alpha_{k,\tau}^1$ can be computed by

$$\alpha_{k,\tau}^1 = \int_{e \in \mathbb{R}} e \cdot \phi_{e_k|\tau}(e)de.$$

For $k = 0$, $\phi_{e_k|\tau}$ is determined by truncating the density function ϕ_{e_0} of the initial state e_0 at all (e, τ), where f_0^0 takes a value of 1 and by normalizing the resulting function, i.e.

$$\phi_{e_0|\tau}(e) = \frac{\phi_{e_0}(e) \cdot (1 - \mathrm{f}_0^0(e, \tau))}{\int_{e \in \mathbb{R}} \phi_{e_0}(e) \cdot (1 - \mathrm{f}_0^0(e, \tau))de}. \tag{19}$$

Since ϕ_{e_0} and f_0^0 are even functions, we conclude that $\phi_{e_0|\tau}$ is even and therefore we have $\alpha_{0,-1}^1 = 0$. Along the same lines, we can show that $\phi_{e_k|k-1}$ is even and $\alpha_{k,k-1}^1 = 0$ for $k \in \{1, \ldots, N-1\}$ by replacing ϕ_{e_0} with ϕ_w in (19). Due to the Bayes rule, the conditional density function $\phi_{e_k|\tau}$ with a fixed τ evolves by the recursion

$$\phi_{e_{k+1}|\tau}(e) = \frac{(\frac{1}{|a|}\phi_{e_k|\tau}(\frac{(\cdot)}{a}) * \phi_w)(e) \cdot (1 - \mathrm{f}_{k+1}^0(e, \tau))}{\int_{e \in \mathbb{R}}(\frac{1}{|a|}\phi_{e_k|\tau}(\frac{(\cdot)}{a}) * \phi_w)(e) \cdot (1 - \mathrm{f}_{k+1}^0(e, \tau))de}.$$

For the prediction step from (8) in the above expression, the scaling refers to the multiplication by a, while the convolution arises from the sum of two independent random variables. The truncation is a result of the incorporation of $\delta_{k+1} = 0$. It can be observed that this recursion preserves symmetry of the conditional density function $\phi_{e_k|\tau}$, as f_k^0 is an even function. Therefore, we have shown that $\alpha^* = 0$ is a fixpoint of Algorithm 1, which completes the proof. ∎

A. Symmetric Unimodal Distributions

A natural question arising from Proposition 4 is whether the person-by-person optimal solution $\alpha^* = 0$ with its optimal event-trigger f^* is also the globally optimal solution. We partly answer this question in the following by assuming that the distributions are unimodal.

Definition 3 (Unimodality): A distribution in $\mathbb{R}$ is called *unimodal*, if there exists $w_0 \in \mathbb{R}$ such that the density function of the distribution $\phi(w)$ is a non-increasing function for $w \geq w_0$ and a non-decreasing function for $w \leq w_0$.

This additional assumption enables us to state the following useful convergence property of Algorithm 1.

Theorem 1: Let $f^* \in \mathcal{F}^{\mathrm{CON}}$ and the initial state e_0 and the noise process $\{w_k\}$ have symmetric and unimodal distributions. Then, $\alpha^* = 0$ is a globally asymptotically stable fixpoint of Algorithm 1.

By considering the evolution of α^i as a dynamical system evolving over variable i, the asymptotic behavior of the iterative Algorithm can be analyzed by means of Lyapunov stability theory and it is shown that $\alpha^* = 0$ is a globally asymptotically stable equilibrium point. The details of the proof can be found in the appendix.

As the iterative Algorithm 1 produces a sequence of pairs (f^i, α^i) whose costs are non-increasing with increasing i, we conclude that 0 is the optimal choice for α, when noise distributions are symmetric and unimodal according to Theorem 1. The optimal state estimator of x_k is then given by the linear predictor in (6) and is therefore independent of the choice of the event-trigger f. The distribution of the initial state x_0 must be also symmetric and unimodal, but its mean $\mathsf{E}[x_0]$ can be chosen arbitrarily. Hence, the symmetry axis of the distribution of x_0 need not to be at zero. In order to determine the optimal $f^* \in \mathcal{F}^{\mathrm{CON}}$, dynamic programming must only be applied once with $\alpha = 0$. Note that such f^* always exists. These results are summarized in the following corollary.

Corollary 1: The optimal event-triggered estimator with $f^* \in \mathcal{F}^{\mathrm{CON}}$ for the first-order system (18) minimizing cost (5) exists and is characterized by (6) and a symmetric threshold policy.

The above corollary is in accordance with [10] and constitutes an alternative way by analyzing the asymptotic behavior of Algorithm 1 to prove that symmetric event-triggering laws are optimal in the presence of symmetric unimodal distributions. Moreover, the iterative algorithm may also

be applied to dynamical systems whose noise distributions do not satisfy these assumptions. Although $\alpha = 0$ is a fix point of the Algorithm 1 by Proposition 4 assuming symmetric density functions, the next section shows that symmetric threshold policies with $\alpha = 0$ can be outperformed by Algorithm 1 by almost 50%. Hence, we can conclude that symmetry of the densities is not sufficient to show that the independent design is optimal. Therefore, additional assumptions are required to show that the independent design is optimal. In the case of Theorem 1 such requirement is given by the unimodality assumption of the density functions.

It is an open question whether Theorem 1 is also valid for higher-order systems. The work in [20] considers the case of multi-dimensional systems for scaled orthonormal system matrices to find structural properties for the joint optimal scheduling and estimation problem. However, neither the use of majorization theory in [10] and [20] nor our iterative approach allow a direct extension to the case of general multi-dimensional systems.

B. Symmetric Bimodal Distributions

This subsection intends to outline the benefits of the iterative algorithm for bimodal noise distributions. We demonstrate numerically how the event-trigger and the estimator can benefit from signaling through the absence of triggering by comparing our solution with the optimal symmetric solution.

Suppose the process (18) with $a = 1$, $\lambda = 0.5$ and the distribution of the initial state, ϕ_{e_0}, and the system noise, ϕ_w, to be given by

$$\phi_{e_0}(\mu, \sigma) = \phi_w(\mu, \sigma) = \frac{1}{2}\phi_{\mathcal{N}}(\mu, \sigma) + \frac{1}{2}\phi_{\mathcal{N}}(-\mu, \sigma)$$

$$\phi_{\mathcal{N}}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

For $\mu = 0$, we retrieve the normal distribution. In the limit $\mu \to 1$, the noise process degrades to a Bernoulli process taking discrete values $\{-1, 1\}$ with probability $\frac{1}{2}$. We select $\mu \in [0, 1)$ and set $\sigma = \sqrt{1 - \mu^2}$ in order to have unit variance for all $\mu \in [0, 1)$. Various density functions for different μ are sketched in Fig. 3a.

We observe that for $\mu < 0.8$ the peaks of the bimodal density function are less distinctive. According to Theorem 1, we can not expect that large gains of the iterative procedure can be attained compared with the optimal symmetric solution for $\mu < 0.8$. A performance comparison of the iterative procedure and the optimal symmetric event-trigger is drawn in Fig. 3b for a
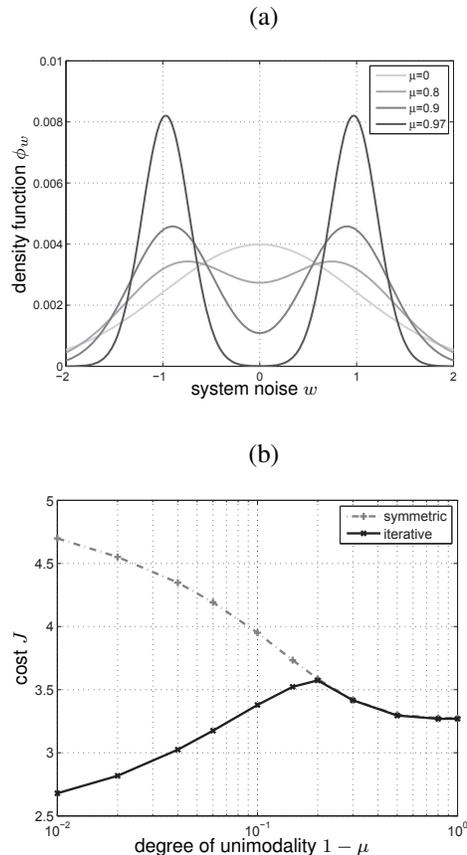
Fig. 3: (a) Density functions for various μ. (b) Performance comparison with varying degree of unimodality drawn on a log. scale.

horizon $N = 10$ and various μ. The initialization for the iterative procedure is chosen to be $\alpha^0 \equiv 0.1$. For $\mu \in [0, 0.8]$, the costs are almost identical as the noise distributions are (approximately) unimodal within this range. This is also anticipated by Theorem 1. For $\mu > 0.8$ a rapid performance improvement can be observed. The value $\mu = 0.8$ can therefore be viewed as the critical border between (approximately) unimodal distributions and density functions with two distinct peaks. In the limit $\mu \to 1$, the costs are reduced by 45% by the iterative procedure compared with the optimal symmetric event-trigger.

Fig. 4 gives an illustrative explanation of the significant performance improvement for $N = 1$ and $\mu = 0.95$. With an initial value $\alpha_0^0 = 0.1$, the iterative algorithm converges to $\alpha_0 \approx 0.95$ and an asymmetric event-trigger $f(x_0) = \mathbb{1}_{[0.25, 1.65]^c}(x_0)$. The event-trigger and estimator have therefore an implicit agreement if no state update is sent over the resource-constrained channel.

In that case, no transmission indicates the estimator that the state x_0 is situated at the right peak resulting in the estimate α_0.
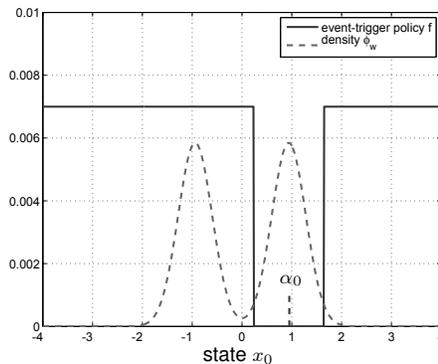

Fig. 4: Asymmetric event-trigger f (scaled by 0.007) resulting from the iterative Algorithm 1 for a bimodal initial distribution φ_w.

VI. SUMMARY

By considering the joint optimal design of state estimator and event-trigger as a two-person problem, we were able to develop an efficient iterative algorithm, which alternates between optimizing the estimator while fixing the event-trigger and vice versa. The iterative method shows special properties in the case of unimodal and symmetric statistics in the uncertainty. In this situation it is shown that the optimal event-triggered estimator can be obtained by a separate design and is given by a linear predictor and a symmetric threshold policy. This result is along previous results and offers an alternative line of proof for showing that such separate design is optimal in case of symmetric unimodal distributions. In the case of symmetric and bimodal distributions, the iterative procedure offers a systematic method, which leads to asymmetric event-triggers and biased estimators that outperform symmetric threshold policies.

Similar properties of the iterative method are likely to hold as well in the case of multi-dimensional systems, but a conclusive derivation for higher-order systems is still an open issue for future research.

APPENDIX

PROOF OF THEOREM 1

Proof: First, we define the following time-variant transformations of e_k and α_{k,τ_k} by

$$y_k = \frac{1}{a^k} e_k, \quad k = 0, \dots, N-1,$$

$$\beta_{k,\tau_k} = \frac{1}{a^k} \alpha_{k,\tau_k}, \; k = 0, \dots, N-1, \; \tau_k = -1, \dots, k-1$$

By this transformation, the running cost and the Bellman operator are defined by

$$\hat{c}_k^{\beta_k}(y_k, \tau_k, \delta_k) = (1 - \delta_k) a^{2k} |y_k - \beta_{k,\tau_k}|^2 + \lambda \delta_k,$$

$$\hat{\mathcal{T}}_k^{\beta_k} \hat{J}_{k+1}(\cdot) = \min_{\delta_k \in \{0,1\}} \hat{c}_k^{\beta_k}(\cdot, \delta_k) + \mathsf{E}\left[\hat{J}_{k+1}(y_{k+1}, \tau_{k+1})|\cdot, \delta_k\right].$$

The optimization problem (16) can then be restated by replacing J with $\hat{J}$ defined by

$$\hat{J}(\mathrm{f}, \beta) = \mathsf{E}^{\mathrm{f}}\left[\sum_{k=0}^{N-1} \hat{c}_k^{\beta_k}(y_k, \tau_k, \delta_k)\right].$$

The event-trigger f_k is a function of y_k and τ_k, where y_k evolves by

$$y_{k+1} = (1 - \delta_k) y_k + v_k, \quad y_0 = e_0.$$

with $v_k = \frac{1}{a^k} w_k$ and the evolution of τ_k is given by (4). It is easy to see that the distribution of v_k is again unimodal and symmetric. In the following, we adapt Algorithm 1 to the transformed system. We consider β^i as a vector in $\mathbb{R}^{\frac{1}{2} N(N+1)}$ that evolves by the procedure defined by (20). By this view, β^i is the state of a non-linear time-invariant discrete-time system described by

$$\mathrm{f}^i = \arg\min_{\mathrm{f}} \hat{J}(\mathrm{f}, \beta^i),$$

$$\beta_{k,\tau}^{i+1} = \mathsf{E}^{\mathrm{f}^i}\left[\sum_{l=\tau}^{k-1} v_l | \delta_{\tau+1} = 0, \dots, \delta_k = 0\right]. \tag{20}$$

In order to analyze the asymptotic behavior with regard to i, we introduce the following Lyapunov candidate $V(\beta^i)$ defined by

$$V(\beta^i) = \|\beta^i\|_\infty.$$

For notational convenience, let β_∞^i be defined as

$$\beta_\infty^i = \|\beta^i\|_\infty.$$

What we want to show first is that for every event-trigger f^i resulting from (20) for a given β^i, we have

$$\mathrm{f}_k^i(\beta_\infty^i + \Delta, \tau) = 0 \quad \Longrightarrow \quad \mathrm{f}_k^i(\beta_\infty^i - \Delta, \tau) = 0,$$
$$\forall \Delta \geq 0, k = 0, \ldots, N - 1, \tau = -1, \ldots, k - 1. \tag{21}$$

The validity of above implication is shown by induction starting with $k = N - 1$. We fix a β^i and apply dynamic programming to obtain f^i. Because of $\hat{J}_N \equiv 0$, the value function $\hat{J}_{N-1}$ is then given by

$$\hat{J}_{N-1}(y, \tau) = \min_{\delta \in \{0,1\}} \hat{c}_{N-1}^{\beta_{N-1}^i}(y, \tau, \delta).$$

Note that the running cost exhibits the symmetry property

$$\hat{c}_k^{\beta_k^i}(\beta_{k,\tau}^i + \Delta, \tau, \delta) = \hat{c}_k^{\beta_k^i}(\beta_{k,\tau}^i - \Delta, \tau, \delta),$$
$$\forall \Delta \in \mathbb{R}, \delta \in \{0, 1\}$$

with $\tau = -1, \ldots, k - 1$ and the monotonicity property

$$0 \leq \Delta_1 \leq \Delta_2$$
$$\Longrightarrow \hat{c}_k^{\beta_k^i}(\beta_{k,\tau}^i + \Delta_1, \tau, \delta) \leq \hat{c}_k^{\beta_k^i}(\beta_{k,\tau}^i + \Delta_2, \tau, \delta)$$

for $\delta \in \{0, 1\}$ and $\tau = -1, \ldots, k - 1$. Both properties are preserved after taking the minimum over δ implying that they are also valid for $\hat{J}_{N-1}$. Therefore, we obtain

$$\hat{J}_{N-1}(\beta_\infty^i + \Delta, \tau) \geq \hat{J}_{N-1}(\beta_\infty^i - \Delta, \tau), \quad \forall \Delta \geq 0 \tag{22}$$

with $\tau = -1, \ldots, N - 1$. For $\Delta \leq \beta_\infty^i - \beta_{k,\tau}^i$, inequality (22) is valid due to the monotonicity property of $\hat{J}_{N-1}$. In case of $\Delta > \beta_\infty^i - \beta_{k,\tau}^i$, we have

$$\hat{J}_{N-1}(\beta_\infty^i - \Delta, \tau)$$
$$= \hat{J}_{N-1}(\beta_\infty^i - \beta_{k,\tau}^i + \beta_{k,\tau}^i - \Delta, \tau)$$
$$= \hat{J}_{N-1}(\beta_{k,\tau}^i + (\beta_{k,\tau}^i - \beta_\infty^i + \Delta, \tau)$$
$$\leq \hat{J}_{N-1}(\beta_\infty^i + \Delta, \tau).$$

The second equality is due to the symmetry property and the inequality is due to the monotonicity property as

$$\beta_{k,\tau}^i \leq \beta_{k,\tau}^i + (\beta_{k,\tau}^i - \beta_\infty^i + \Delta) \leq \beta_\infty^i + \Delta.$$

By knowing that the value function $\hat{J}_{N-1} = \lambda$ is constant for all pairs (y, τ), when $\delta_{N-1} = 1$, we have

$$\mathrm{f}_{N-1}^i(\beta_\infty^i - \Delta, \tau) = 1$$

$$\implies \lambda = \hat{J}_{N-1}(\beta_\infty^i - \Delta, \tau) \leq \hat{J}_{N-1}(\beta_\infty^i + \Delta, \tau)$$

$$\implies J_{N-1}(\beta_\infty^i + \Delta, \tau) = \lambda$$

$$\implies \mathrm{f}_{N-1}^i(\beta_\infty^i + \Delta, \tau) = 1.$$

Next, we show that by applying the Bellman operator will preserve the inequality given by (22). Assume, we have

$$\hat{J}_{k+1}(\beta_\infty^i + \Delta, \tau) \geq \hat{J}_{k+1}(\beta_\infty^i - \Delta, \tau), \quad \forall \Delta \geq 0 \tag{23}$$

with $\tau = -1, \ldots, k-1$. We want to show statement (23) implies

$$\hat{J}_k(\beta_\infty^i + \Delta, \tau) \geq \hat{J}_k(\beta_\infty^i - \Delta, \tau), \quad \forall \Delta \geq 0 \tag{24}$$

with $\tau = -1, \ldots, k-1$. The Bellman equation is

$$\hat{J}_k = \hat{\mathcal{T}}_k^{\beta_k^i} \hat{J}_{k+1}.$$

For all pairs (y, τ), where the argument of the minimization in $\hat{\mathcal{T}}_k^{\beta_k^i}$ yields $\delta_k = 1$, $\hat{J}_k$ is constant. This also implies that $\hat{J}_k$ takes its maximum for these pairs. In the following, we are interested in outcomes for $\hat{J}_k$ in case of $\delta_k = 0$. Along the same lines as for $\hat{J}_{N-1}$, we obtain for the running cost $\hat{c}_k^{\beta_k^i}$

$$\hat{c}_k^{\beta_k^i}(\beta_\infty^i + \Delta, \tau, \delta) \geq \hat{c}_k^{\beta_k^i}(\beta_\infty^i - \Delta, \tau, \delta), \qquad \forall \Delta \in \mathbb{R}, \delta \in \{0, 1\} \tag{25}$$

with $\tau = -1, \ldots, k-1$. We rewrite $\hat{J}_{k+1}$ to

$$\hat{J}_{k+1} = \hat{J}_{k+1}^{\text{SYM}} + \hat{J}_{k+1}^{\text{REM}}$$

with

$$\hat{J}_{k+1}^{\text{SYM}}(y, \tau) = \begin{cases} \hat{J}_{k+1}(y, \tau), & y \leq \beta_\infty^i, \\ \hat{J}_{k+1}(\beta_\infty^i + (\beta_\infty^i - y), \tau), & y > \beta_\infty^i, \end{cases} \tag{26}$$

$$\hat{J}_{k+1}^{\text{REM}}(y, \tau) = J_{k+1}(y, \tau) - \hat{J}_{k+1}^{\text{SYM}}(y, \tau). \tag{27}$$

By the assumption (23), we have

$$\hat{J}_{k+1}^{\mathrm{REM}}(y,\tau) \begin{cases} = 0, & y \leq \beta_\infty^i, \\ \geq 0, & y > \beta_\infty^i. \end{cases} \tag{28}$$

Taking the expectation of $\hat{J}_{k+1}$ given δ_k, y_k and τ_k, gives either a constant function over (y_k, τ_k) for $\delta_k = 1$ or is given by convolution with the density function of v_k for $\delta_k = 0$ denoted by ϕ. By assumption the density function ϕ is symmetric and unimodal. By linearity of the convolution operator, we follow

$$\mathsf{E}\left[\hat{J}_{k+1}|\cdot, \tau_k, \delta_k = 1\right] = J_{k+1}^{\mathrm{SYM}}(\cdot, \tau_k) * \phi + J_{k+1}^{\mathrm{REM}}(\cdot, \tau_k) * \phi. \tag{29}$$

For the first term of (29), we see that symmetry is preserved, i.e.,

$$(J_{k+1}^{\mathrm{SYM}}(\cdot, \tau_k) * \phi)(\beta_\infty^i + \Delta) = (J_{k+1}^{\mathrm{SYM}}(\cdot, \tau_k) * \phi)(\beta_\infty^i - \Delta) \tag{30}$$

for $\Delta \in \mathbb{R}$. On the other hand due to (28) and

$$\phi(y - (\beta_\infty^i + \Delta)) \geq \phi(y - (\beta_\infty^i - \Delta)), \Delta \geq 0, y \geq \beta_\infty^i,$$

we have for any $\Delta \geq 0$

$$(J_{k+1}^{\mathrm{REM}}(\cdot, \tau_k) * \phi)(\beta_\infty^i + \Delta) \geq (J_{k+1}^{\mathrm{REM}}(\cdot, \tau_k) * \phi)(\beta_\infty^i - \Delta). \tag{31}$$

Summing up the terms and taking the minimum to obtain $\hat{J}_k$, we obtain statement (24) by using (25), (30) and (31). By induction, statement (24) is valid for all $k \in \{0, \ldots, N-1\}$. Along the same lines as for $N-1$, we follow (21) from the statement in (24). Equivalently to (21), it can be showed that

$$\mathrm{f}_k^i(-\beta_\infty^i - \Delta, \tau) = 0 \implies \mathrm{f}_k^i(-\beta_\infty^i + \Delta, \tau) = 0,$$

$$\forall \Delta \geq 0, k = 0, \ldots, N-1, \tau = -1, \ldots, k-1.$$

Let $\phi_{y_k|\tau}^i$ be defined as the density function of the conditional probability distribution of y_k given τ_k and $\delta_k = 0$, when using event-trigger f^i. Then, $\beta_{k,\tau}^{i+1}$ is given by

$$\beta_{k,\tau}^{i+1} = \int_{y \in \mathbb{R}} y \cdot \phi_{y_k|\tau}^i(y) dy.$$

By assuming f^i satisfying (21), we show inductively that

$$\phi^i_{y_k|\tau}(\beta^i_\infty + \Delta) \le \phi^i_{y_k|\tau}(\beta^i_\infty - \Delta), \quad \forall \Delta \ge 0,$$

$$k = 0, \ldots, N-1, \tau = -1, \ldots, k-1.$$

(32)

For $k = 0$, $\phi^i_{y_k|\tau}$ is calculated by truncating the density function ϕ_{y_0} of the initial state y_0 at all (y, τ), where f^i_k takes a value of 1 and by normalizing the resulting function, i.e.

$$\phi^i_{y_0|\tau}(y) = \frac{\phi_{y_0}(y) \cdot (1 - f^i_0(y, \tau))}{\int_{y \in \mathbb{R}} \phi_{y_0}(y) \cdot (1 - f^i_0(y, \tau)) dy}.$$

As ϕ_{y_0} is an even and unimodal function, we have

$$\phi^i_{y_0|\tau}(\beta^i_\infty + \Delta) \le \phi^i_{y_0|\tau}(\beta^i_\infty - \Delta),$$

$$\Delta \ge 0, f^i_k(\beta^i_\infty + \Delta, \tau) = 0.$$

For all (y, τ) with $f^i_k(\beta^i_\infty + \Delta, \tau) = 1$, we have

$$\phi^i_{y_0|\tau}(\beta^i_\infty + \Delta) = 0,$$

which trivially validates inequality (32). Similarly as for $k = 0$ and $\tau = -1$, we can prove the validity of (32) for $k \in \{1, \ldots, N-1\}$ and $\tau = k-1$ by replacing the density function ϕ_{y_0} by the density function $\phi_{v_{k-1}}$ of the noise variable v_{k-1}. By assuming that inequality (32) is satisfied for time step k, we prove that (32) holds for $k+1$ for an arbitrary $k \in \{0, \ldots, N-2\}$ and fixed $\tau \in \{-1, \ldots, k-1\}$. For a fixed τ, $\phi^i_{y_k|\tau}(y)$ arises from the recursion

$$\phi^i_{y_{k+1}|\tau}(y) = \frac{(\phi^i_{y_k|\tau} * \phi_{v_k})(y) \cdot (1 - f^i_{k+1}(y, \tau))}{\int_{y \in \mathbb{R}} (\phi^i_{y_k|\tau} * \phi_{v_k})(y) \cdot (1 - f^i_{k+1}(y, \tau)) dy}.$$

(33)

As having already been observed for $\hat{J}_{k+1}$, the convolution of $\phi^i_{y_k|\tau}$ with ϕ_{v_k} preserves the inequality (32). With the same arguments as for $k = 0$, we follow that

$$\phi^i_{y_k|\tau}(\beta^i_\infty + \Delta) \le \phi^i_{y_k|\tau}(\beta^i_\infty - \Delta), \qquad \Delta \ge 0$$

implies

$$\phi^i_{y_{k+1}|\tau}(\beta^i_\infty + \Delta) \le \phi^i_{y_{k+1}|\tau}(\beta^i_\infty - \Delta), \qquad \Delta \ge 0,$$

which concludes the induction.

Inequality (32) implies that $\beta^{i+1}_{k,\tau} \le \beta^i_\infty$. Similarly, it can be showed that $\beta^{i+1}_{k,\tau} \ge -\beta^i_\infty$. In fact, it is straight forward to see that the inequalities are strict for all $\beta^i_\infty \ne 0$ and therefore the

Lyapunov candidate V decreases with increasing i for all $\beta \neq 0$. Hence, the iterative procedure defined in (20) converges to 0 for any initial condition of β. By transforming β back into the initial state space system, we can conclude the proof. ∎

REFERENCES

[1] K. Åström and B. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," in *41th IEEE Conference on Decision and Control (CDC)*, 2002.

[2] Y. Xu and J. Hespanha, "Optimal communication logics in networked control systems," in *43rd IEEE Conference on Decision and Control (CDC)*, vol. 4, 2004, pp. 3527–3532.

[3] M. Rabi, G. V. Moustakides, and J. S. Baras, "Adaptive sampling for linear state estimation," *SIAM Journal on Control and Optimization*, vol. 50, no. 2, pp. 672–702, 2012.

[4] J. F. 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.

[5] A. Molin and S. Hirche, "On the optimality of certainty equivalence for event-triggered control systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 470–474, 2013.

[6] ——, "Price-based adaptive scheduling in multi-loop control systems with resource constraints," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3282–3295, 2014.

[7] 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.

[8] S. Thrun, M. Montemerlo, D. Koller, B. Wegbreit, J. Nieto, and E. Nebot, "FastSLAM: An efficient solution to the simultaneous localization and mapping problem with unknown data association," *Journal of Machine Learning Research*, vol. 4, no. 3, pp. 380–407, 2004.

[9] J. Sijs, B. Noack, and U. D. Hanebeck, "Event-based state estimation with negative information," in *16th International Conference on Information Fusion (FUSION)*. IEEE, 2013, pp. 2192–2199.

[10] G. Lipsa and N. Martins, "Remote state estimation with communication costs for first-order LTI systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, 2011.

[11] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*. Cambridge university press, 1952.

[12] B. Hajek, K. Mitzel, and S. Yang, "Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm," *IEEE Transactions on Information Theory*, vol. 54, no. 2, pp. 608–622, 2008.

[13] J. Karlsson, A. Gattami, T. J. Oechtering, and M. Skoglund, "Iterative source-channel coding approach to Witsenhausen's counterexample," in *American Control Conference (ACC)*, 2011, pp. 5348 –5353.

[14] J. Karlsson and M. Skoglund, "Optimized low-delay source-channel-relay mappings," *IEEE Transactions on Communications*, vol. 58, no. 5, pp. 1397–1404, 2010.

[15] L. Bao, M. Skoglund, and K. H. Johansson, "Encoder–decoder design for event-triggered feedback control over bandlimited channels," in *American Control Conference (ACC)*. IEEE, 2006, pp. 4183–4188.

[16] S. Park and N. C. Martins, "Individually optimal solutions to a remote state estimation problem with communication costs," in *53rd IEEE Conference on Decision and Control (CDC)*, 2014, pp. 4014–4019.

[17] Q. Du, V. Faber, and M. Gunzburger, "Centroidal voronoi tessellations: applications and algorithms," *SIAM review*, vol. 41, no. 4, pp. 637–676, 1999.

[18] M. M. Vasconcelos and N. C. Martins, "Estimation over the collision channel: Structural results," in *51st Annual Allerton Conference on Communication, Control, and Computing*.   IEEE, 2013, pp. 1114–1119.

[19] J. Chakravorty and A. Mahajan, "Distortion-transmission trade-off in real-time transmission of markov sources," *arXiv preprint arXiv:1412.3199*, 2014.

[20] A. Nayyar, T. Basar, D. Teneketzis, and V. V. Veeravalli, "Optimal strategies for communication and remote estimation with an energy harvesting sensor," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2246–2260, 2013.

[21] A. Molin and S. Hirche, "An iterative algorithm for optimal event-triggered estimation," in *4th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS)*, 2012, pp. 64–69.

[22] D. P. Bertsekas, *Dynamic programming and optimal control. Vol. I.*   3rd ed., Athena Scientific, Belmont, MA, 2005.