Event-Based State Estimation with Variance-Based Triggering

Sebastian Trimpe, *Member, IEEE*, and Raffaello D'Andrea, *Fellow, IEEE*

***Abstract*—An event-based state estimation scenario is considered where multiple distributed sensors sporadically transmit observations of a linear process to a time-varying Kalman filter via a common bus. The triggering decision is based on the estimation variance: each sensor runs a copy of the Kalman filter and transmits its measurement only if the associated measurement prediction variance exceeds a tolerable threshold. The resulting variance iteration is a new type of Riccati equation, with switching between modes that correspond to the available measurements and depend on the variance at the previous step. Convergence of the switching Riccati equation to periodic solutions is observed in simulations, and proven for the case of an unstable scalar system (under certain assumptions). The proposed method can be implemented in two different ways: as an event-based scheme where transmit decisions are made online, or as a time-based periodic transmit schedule if a periodic solution to the switching Riccati equation is found.**

***Index Terms*—Event-based state estimation, distributed estimation, sensor scheduling, networked control systems, switching Riccati equation, periodic solution.**

I. Introduction

Novel control strategies and improvements in sensor, actuator and network technology will allow the next generation of control systems to tightly integrate the physical world with computation and communication. Referred to as cyber-physical systems (CPSs) [1], these highly integrated systems will extend present-day networked systems (such as networked control systems (NCSs) [2] and wireless sensor networks (WSNs) [3]) in both size and complexity.

As the number of interconnected entities in future CPSs increases, the cost of communication will become a significant factor. Communication is costly even in today's networked systems. In NCSs, where a multi-purpose communication network is shared by many different control, sensor and actuator units, a sensor node cannot transmit its measurement without preventing other units from using the network or causing load-induced delays. In WSNs, the transmission of a sensor measurement to a remote estimator consumes energy that is often a significant fraction of the system's overall energy balance.

While the future of CPSs in areas such as transportation, power systems, smart buildings, mobile robots and process plants is promising, the cost of communication must be managed if CPSs are to meet their potential. It will be vital, for example, to consider the communication network as a shared resource, and to design the control and estimation algorithms in tandem with the network access strategy.

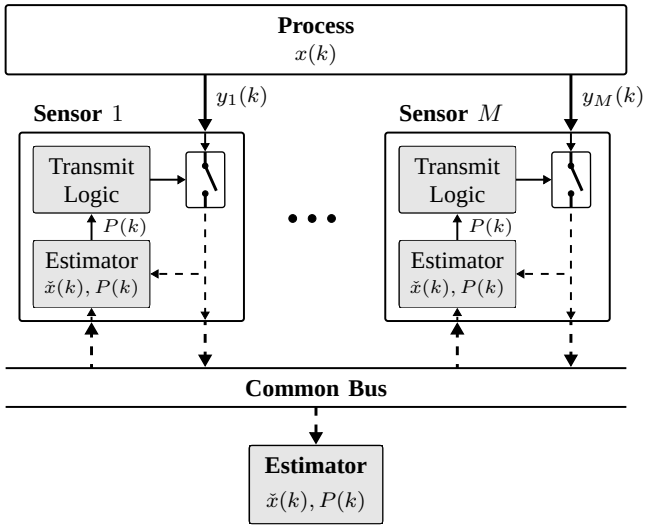


Fig. 1. Distributed event-based state estimation problem. The state $x(k)$ of a linear process is observed by M sensor agents, which sporadically transmit their measurements $y_j(k)$ over a common bus. (Solid lines denote continuous flow of data, dashed lines indicate event-based communication, and communication is assumed without delay and data loss.) Estimator nodes connected to the bus receive the measurements and keep track of the conditional state mean $\check{x}(k)$ and variance $P(k)$. Each sensor makes the decision whether to transmit its local measurement based on the estimation variance $P(k)$, thus linking the transmit decision to the estimation performance. The gray blocks constitute the event-based state estimator to be designed herein. The depicted scheme can be applied in different scenarios where communication is costly. In a monitoring application, a remote estimator centrally fuses all sensor data received from the bus (as shown here). In a networked control system, where the agents are also equipped with actuation, the state estimates can be used locally for feedback control.

This article considers the problem of estimating the state of a dynamic system from multiple distributed sensors in a scenario where the communication of sensor measurements is costly. We propose a method where data is transmitted only when certain events indicate that the data is required to meet constraints on the estimator performance (expressed as tolerable bounds on the error variance). Thus the transmit decision is linked to its contribution to the estimator performance and data is exchanged only when needed.

Figure 1 depicts the distributed state estimation problem and the event-based communication strategy used to address it. The key idea is that each agent transmits its local sensor measurement *only if it is required in order to meet a certain estimation performance*. To be able to make this decision, each

S. Trimpe is with the Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany (e-mail: strimpe@tuebingen.mpg.de).

R. D'Andrea is with the Institute for Dynamic Systems and Control, ETH Zurich, 8092 Zurich, Switzerland (e-mail: rdandrea@ethz.ch).

This work was supported by the Swiss National Science Foundation (SNSF).

Accepted final version. Accepted for publication in: IEEE Transactions on Automatic Control, 2014.

agent implements a state estimator that is connected to the common bus. Since the state estimate is computed based on data received over the bus only (the local sensor data is used only when also broadcast) and since we assume a loss and delay-free network, the estimates are the same on all agents and represent the common information in the network. The estimator can hence be used to make the transmit decision: if the other agents' estimate of a particular measurement is already "good enough," it is not necessary to communicate this measurement; if the common estimate is poor, on the other hand, the measurement is transmitted so that all agents can update their estimates. The estimators are implemented as time-varying Kalman filters, which compute the mean and variance of the state conditioned on the received measurements.

Different decision rules for determining whether an estimate is "good enough" are conceivable. In [4], [5], for example, a constant threshold logic on the difference between the actual measurement and its prediction is used. Herein, we consider a different approach where the decision is based on the variance: a measurement is broadcast if its prediction variance exceeds a tolerable bound, which indicates that the uncertainty when predicting the measurement is too large. If a transmission is triggered by a condition on the estimation variance, we refer to this as *variance-based triggering*.

As opposed to making the transmit decision based on real-time measurement data (where, for a stochastic process, the transmit decision is a random variable), the approach herein permits an off-line analysis of the (deterministic) estimation variance iteration, provided that the observed process is stationary and its statistics are known in advance. Specifically, if a periodic solution of the variance iteration is found, it corresponds to a periodic sending sequence for each sensor, and hence allows for a straightforward implementation of the resulting communication logic. The periodic sending sequence can be computed in advance and fixed for each sensor. Such a time-based implementation represents an alternative to the event-based implementation shown in Fig. 1 where transmit decision are made on-line. The proposed method can hence also be used as a tool for off-line design of periodic sensor schedules. The recursive equation for the estimator variance is a switching equation that represents a new type of Riccati equation. A focus of this article is on studying the convergence of this equation for the special case of a scalar unstable system.

A. Contributions and Outline

The main contributions of this article are the following:

- Variance-based triggering (combined with a Kalman filter) is proposed as a novel framework for event-based state estimation.
- A new type of switching Riccati equation is obtained as the variance iteration for this event-based estimator.
- The convergence properties of the switching Riccati equation are studied for an unstable scalar process, and a sufficiency result for asymptotic periodicity is derived (Theorem 2): if two assumptions are satisfied, global convergence of the Riccati equation to a periodic solution is guaranteed.

We first mentioned the idea of event-based state estimation with variance-based triggering in the conference paper [6], where we applied the method on the NCS of the Balancing Cube [7]. By using the event-based estimator for feedback control, we achieved a significant reduction in average communication at only a mild decrease in control performance. The experimental results from [6] are not repeated herein. This article includes the proofs for the convergence result of the scalar switching Riccati equation, which were omitted in a preliminary version of the result in [8].

This article continues as follows: After a review of related work and introduction of notation in the next subsections, the event-based state estimator and its corresponding Riccati equation are derived in Sec. II. In Sec. III, we illustrate the periodic behavior of the switching Riccati equation through simulation examples of a scalar and a multivariable process. The convergence result for the scalar case is derived in Sec. IV, and we conclude with a discussion of the results in Sec. V.

B. Related Work

Event-based strategies are a popular means of ensuring efficient use of the communication resource in NCSs or CPSs (see [9] and references therein). As opposed to traditional time-triggered transmission of data, event-based approaches transmit data only when required to meet a certain specification of the control system (e.g. closed-loop stability, control or estimator performance). Event-based state estimation problems with a single sensor and a single estimator node have been studied in [9]–[16], for example. Event-based state estimation problems for distributed or multi-agent systems have been looked at in [4], [5], [17].

The basic idea of implementing state estimators on the agents of an NCS in order to reduce communication of sensor data was first presented in [18]. Therein, each agent uses a model to predict the other agents' measurements at times when these are not transmitted (because the prediction error is below a threshold), and the agent *resets* parts of the state vector when new measurement data becomes available. In contrast, the Kalman filters used herein *fuse* model-based predictions with the received measurements. Communication schemes like these where, in order to reduce network traffic, sensor data is not sent at every time step, are also referred to as *controlled communication*, [2], [13], [19].

In most of the above-mentioned references for the single sensor/single estimator case, the sensor node transmits a local state estimate (obtained from a Kalman filter on the sensor) to the remote estimator, rather than the raw measurement. While this seems to be the method of choice for the single sensor agent case (the local state estimate contains the fused information of all past measurements), communicating raw measurements has a practical advantage for the case of multiple agents with coupled dynamics. For an agent to fuse another agent's measurement with its local state estimate, it must know the variance of the measurement conditioned on the state. This is usually known in form of a sensor model. To optimally fuse another agent's state estimate (with coupled dynamics), on the other hand, the variance associated with

the estimate would have to be known. Since this variance is, however, only known to the agent that generated the estimate, it would have to be communicated over the network as well, hence, increasing the network load. The method herein makes no assumptions on the dynamic coupling between the system parts that are observed by the various sensors.

In the above-mentioned references on event-based estimation, an event is triggered by some condition on real-time data (measurement or state); that is, in a stochastic framework, data transmission is a random event. In contrast, the variance-based trigger used herein depends on the prediction variance at the previous step. The resulting variance iteration is deterministic and depends on the problem data only. A condition on the variance to trigger sensor transmissions is considered in [20] in a slightly different framework. Therein, the authors consider two heterogeneous sensors: a condition on the estimator variance is used to decide which of the sensors will transmit its measurement to a remote estimator at any given time step. Whereas the average communication rate is constant in [20], we seek to reduce the average sensor transmission rate, and to have the option to not transmit any data at a time step. The authors in [20] also observe convergence of the estimator variance to periodic sequences in their scenario, but they do not prove this convergence.

As mentioned previously, the event-based estimation method herein can also be used for off-line design of sensor schedules (when a periodic solution to the Riccati equation is implemented as a time-based schedule). Related sensor scheduling problems are discussed in [21], [22], and, more recently, [23]. Meier et al. [21] and Kushner [22] address a finite horizon optimal scheduling problem, where m sensor observations are to be scheduled over a finite horizon $T > m$ to minimize some cost function. They show that the optimal schedule can be determined off-line (as is the case herein), and Kushner finds that the m observation instants tend to be equally spaced (i.e. "periodic" over the finite horizon) for unstable scalar systems and a large time horizon T. Zhang et al. [23] considered an infinite horizon optimal scheduling problem, for which they established that the optimal sensor schedule can be approximated arbitrarily well by a periodic one. In contrast to the aforementioned works, we do not cast the reduced communication estimation problem as an optimization problem. Instead of fixing the number of transmissions a priori, we start from specifications of the desired estimation quality (expressed as bounds on the measurement prediction variance) and use event-triggers to determine whether a transmission needs to happen or communication can be saved. Notably, periodic schedules appear herein, as well as in [20], [22], [23], as limiting cases or suboptimal solutions despite different underlying problem formulations.

The Riccati iteration obtained for the event-based estimation problem herein is related to Riccati equations of other well-known Kalman filtering problems with different sensor transmission policies. For full communication (all sensors transmit at every time step), the discrete-time Riccati equation for the standard Kalman filtering problem is recovered, [24]. When periodic transmission schedules are fixed a priori, the problem can be cast as a linear periodic system, and the estimation variance is thus captured by the discrete-time periodic Riccati equation, [25]. Note that the problem considered herein is different in that we do not a priori assume a periodic transmit sequence, but we show that a periodic sequence results from the event-based estimation problem. In Kalman filtering with intermittent observations [26], measurement arrival at the filter is subject to random data loss modeled as a Bernoulli process, and the filter variance becomes a random variable itself. In contrast to being governed by an external (random) process, measurement transmissions herein are triggered "internally" by the estimator whenever new data is needed.

Event-based state estimation is closely related to event-based control. In fact, an event-based controller is obtained when the output of an event-based estimator is connected to a state-feedback controller (as is done in the experimental applications in [4]–[6]). Results in [27] and [28] suggest this structure for event-based control. Therein, the authors prove that the combination of an optimal state-feedback controller (designed using standard methods) with an optimal event-based state estimator (from a joint design of estimator and transmission logic) yields the optimal event-based controller (for linear systems with a centralized controller and a quadratic finite-horizon cost).

C. Notation and Preliminaries

We use $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{N}$, and $\mathbb{N}^+$ to denote real numbers, integers, nonnegative integers, and positive integers, respectively. By $\mathrm{E}\left[\cdot|\cdot\right]$ and $\mathrm{Var}\left[\cdot|\cdot\right]$, we denote the conditional expected value and the conditional variance. A normally distributed random variable z with mean m and covariance matrix V is denoted by $z \sim \mathcal{N}(m, V)$.

For $i, j \in \mathbb{Z}$ and $N \in \mathbb{N}^+$, we define the binary operator '$-_N$' as follows:

$$i -_N j = \begin{cases} \mathrm{mod}(i-j, N) & \text{if } \mathrm{mod}(i-j, N) > 0 \\ N & \text{if } \mathrm{mod}(i-j, N) = 0, \end{cases} \tag{1}$$

where $\mathrm{mod}(i, N) \in \{0, \dots, N-1\}$ is the (nonnegative) remainder of i divided by N. Hence, '$-_N$' is the subtraction with subsequent modulo N operation, except that a resulting 0 is replaced by N.

For a symmetric matrix $X \in \mathbb{R}^{n\times n}$, we write $X > 0$ and $X \geq 0$ to mean that X is positive definite and positive semi-definite, respectively. For a matrix $A \in \mathbb{R}^{m\times n}$, $A_{j:}$ denotes the jth row, and $[A_{j:}]_{j\in J}$ with $J \subseteq \{1, \dots, m\}$ denotes the matrix constructed from stacking the rows $A_{j:}$ for all $j \in J$. Further, $\mathrm{diag}[A_{jj}]_{j\in J}$ denotes the diagonal matrix with entries A_{jj}, $j \in J$, on its diagonal.

We define the binary indicator function $\mathbf{1}_X$ such that $\mathbf{1}_X = 1$ if statement X is true, and $\mathbf{1}_X = 0$ otherwise.

Consider the iteration

$$p(k+1) = h(p(k)), \quad p(0) = p_0 \geq 0, \tag{2}$$

with a function $h : D \to \mathbb{R}^n$, $D \subseteq \mathbb{R}^n$. For h being applied m times, we write h^m; that is, for $m \in \mathbb{N}$,

$$p(k+m) = h^m(p(k)) = \underbrace{h(h(\dots(h}_{m}(p(k))\dots))), \tag{3}$$

where $h^0(p(k)) := p(k)$. For the domain of a function h, we write $\mathrm{dom}(h)$. We use the following definitions to characterize periodic solutions of (2):

Definition 1 (adapted from [29]): Let $p^* \in \mathrm{dom}(h)$. Then p^* is called an *N-periodic point* of (2) if it is a fixed point of h^N, that is, if

$$h^N(p^*) = p^*. \tag{4}$$

The periodic orbit of p^*, $\{p^*, h(p^*), h^2(p^*), \ldots, h^{N-1}(p^*)\}$, is called an *$N$-cycle*, and N is called the *period*.

Definition 2: A solution to (2) is called *asymptotically N-periodic* for the initial condition $p(0) = p_0$ if

$$\lim_{m\to\infty} h^{mN}(p_0) = p^*, \tag{5}$$

where p^* is an N-periodic point of (2).

For a function h and collections of intervals $\mathcal{I}_1$ and $\mathcal{I}_2$, we write $\mathcal{I}_1 \xrightarrow{h} \mathcal{I}_2$ to indicate that each interval from $\mathcal{I}_1$, when mapped by h, is contained in an interval in $\mathcal{I}_2$; that is,

$$\mathcal{I}_1 \xrightarrow{h} \mathcal{I}_2 \quad \Leftrightarrow \quad \forall I_1 \in \mathcal{I}_1, \exists I_2 \in \mathcal{I}_2 \,:\, h(I_1) \subseteq I_2. \tag{6}$$

II. Event-Based State Estimator

We consider the stochastic linear time-invariant system

$$x(k) = A\,x(k-1) + v(k-1) \tag{7}$$
$$y(k) = C\,x(k) + w(k), \tag{8}$$

where k is the discrete time index, $x(k) \in \mathbb{R}^n$ is the state, $y(k) \in \mathbb{R}^M$ its observation (M the number of sensors), and all matrices are of corresponding dimensions. The process noise, the measurement noise, and the initial state $x(0)$ are assumed mutually independent, normally distributed with $v(k) \sim \mathcal{N}(0, Q)$, $w(k) \sim \mathcal{N}(0, R)$, $x(0) \sim \mathcal{N}(x_0, P_0)$, $Q \geq 0$, $R > 0$, and $P_0 \geq 0$. We assume that (A, C) is detectable (i.e. the process is detectable when measurements from all sensors are combined, but not necessarily from an individual sensor), (A, Q) is stabilizable, and R is diagonal. The latter assumption means that the measurement noise is mutually independent for any two sensors considered, which is often the case in practice. The presented state estimation method can, however, be readily extended to the case of block diagonal R by sending blocks of correlated measurements at once.

Remark 1: For ease of notation, we consider an unforced system; that is, no input $u(k-1)$ in (7). Provided the inputs are known by all sensor agents at all times (such as when they represent a known reference signal or when they are shared over the network), the extension of the event-based state estimator to this case is straightforward, and the analysis of the estimator variance in the following remains unchanged.

We seek an algorithm to recursively compute an estimate of the state $x(k)$ from measurements received up to time k, and the problem parameters given by (A, C, Q, R, P_0). If the full measurement vector $y(k)$ is available at time k, the problem is solved by the standard Kalman filter (Sec. II-A). In Sec. II-B, we present the event-based state estimator, which consists of a Kalman filter that uses a reduced set of measurements as an input and a rule for deciding whether to transmit a measurement.

A. Full Communication Kalman Filter

It is well known that the Kalman filter is the optimal Bayesian state estimator for the process (7), (8) because it keeps track of the Gaussian conditional probability distribution of the state $x(k)$ conditioned on all measurements up to time k, $\mathcal{Y}(k) := \{y(1), \ldots, y(k)\}$ (see [24], for example). To distinguish this Kalman filter from the event-based filter derived below, we refer to it as the *full communication Kalman filter*. Under the above assumptions, the state prediction variance $\mathrm{Var}\,[x(k)|\mathcal{Y}(k-1)]$ converges to $\bar{P} > 0$, which is the unique positive semidefinite solution to the discrete algebraic Riccati equation (DARE):

$$\bar{P} = A\bar{P}A^{\mathrm{T}} + Q - A\bar{P}C^{\mathrm{T}}(C\bar{P}C^{\mathrm{T}} + R)^{-1}C\bar{P}A^{\mathrm{T}}. \tag{9}$$

We write $\bar{P} = \mathrm{DARE}(A, C, Q, R)$.

B. Event-Based Kalman Filter

Denote by $J(k) \subseteq \{1, \ldots, M\}$ the subset of sensors that transmit their measurement at time k. We make precise how we choose $J(k)$ later in this section. Since communication is assumed to be instantaneous and without data loss, $J(k)$ is also the set of measurements available at the estimator at time k. The corresponding measurement equation is then given by

$$\tilde{y}(k) = \tilde{C}(k)\,x(k) + \tilde{w}(k), \tag{10}$$

where $\tilde{y}(k) = [y_j(k)]_{j\in J(k)}$ is the vector of those measurements available at time k, $\tilde{w}(k) \sim \mathcal{N}(0, \tilde{R}(k))$, and output and measurement noise variance matrices are constructed as

$$\tilde{C}(k) = [C_{j:}]_{j\in J(k)}, \quad \tilde{R}(k) = \mathrm{diag}[R_{jj}]_{j\in J(k)}. \tag{11}$$

Notice that $\tilde{y}(k) \in \mathbb{R}^{m(k)}$, $\tilde{w}(k) \in \mathbb{R}^{m(k)}$, $\tilde{C}(k) \in \mathbb{R}^{m(k)\times n}$, and $\tilde{R}(k) \in \mathbb{R}^{m(k)\times m(k)}$ have time varying dimensions with $m(k) \leq M$. This includes the case $m(k) = 0$; that is, at time k there is no measurement available at the estimator. In order to avoid special treatment of this case, we use the convention that the measurement update step in the Kalman filter below is omitted in case $m(k) = 0$.

For any given sequence of $\tilde{C}(k)$ and $\tilde{R}(k)$, the distribution of the state $x(k)$ conditioned on the set of available measurements $\tilde{\mathcal{Y}}(k) = \{\tilde{y}(l) \mid 0 \leq l \leq k\}$ is Gaussian, [24]. The Kalman filter keeps track of the conditional means and variances, $\check{x}(k|k-1) = \mathrm{E}\,[x(k)|\tilde{\mathcal{Y}}(k-1)]$, $\check{x}(k|k) = \mathrm{E}\,[x(k)|\tilde{\mathcal{Y}}(k)]$, $\check{P}(k|k{-}1) = \mathrm{Var}\,[x(k)|\mathcal{Y}(k{-}1)]$, and $\check{P}(k|k) = \mathrm{Var}\,[x(k)|\tilde{\mathcal{Y}}(k)]$. The filter equations are

$$\check{x}(k|k-1) = A\check{x}(k-1|k-1) \tag{12}$$
$$\check{P}(k|k-1) = A\check{P}(k-1|k-1)A^{\mathrm{T}} + Q \tag{13}$$
$$\check{K}(k) = \check{P}(k|k-1)\,\tilde{C}^{\mathrm{T}}(k) \cdot \big(\tilde{C}(k)\check{P}(k|k-1)\tilde{C}^{\mathrm{T}}(k) + \tilde{R}(k)\big)^{-1} \tag{14}$$
$$\check{x}(k|k) = \check{x}(k|k-1) + \check{K}(k)\big(\tilde{y}(k) - \tilde{C}(k)\check{x}(k|k-1)\big) \tag{15}$$
$$\check{P}(k|k) = \big(I - \check{K}(k)\tilde{C}(k)\big)\check{P}(k|k-1). \tag{16}$$

The filter is initialized with $\check{x}(0|0) = x_0$ and $\check{P}(0|0) = P_0$.

For notational convenience, we use $P(k) := \check{P}(k|k-1)$ for the state prediction variance. The prediction variance captures

the uncertainty about $x(k)$ given all measurements up to the previous time step $k-1$. Similarly, $\mathrm{Var}\,[y_j(k)|\tilde{\mathcal{Y}}(k-1)] = C_j P(k) C_j^{\mathrm{T}} + R_{jj}$ captures the uncertainty in predicting the measurement $y_j(k)$. A measurement $y_j(k)$ is transmitted and used to update the estimator if, and only if, its prediction variance exceeds a tolerable bound. Since the event-based Kalman filter cannot do better than the full communication filter, we use a threshold δ on the difference $\mathrm{Var}\,[y(k)|\tilde{\mathcal{Y}}(k-1)] - \lim_{k\to\infty} \mathrm{Var}\,[y(k)|\mathcal{Y}(k-1)] = C_j(P(k)-\bar{P})C_j^{\mathrm{T}}$ for the transmit decision. Hence, we use the transmit rule

$$\text{transmit } y_j(k) \Leftrightarrow\ C_j\big(P(k)-\bar{P}\big)C_j^{\mathrm{T}} \geq \delta_j \qquad (17)$$

for sensor j, where the design parameter δ_j captures the tolerable deviation of the jth sensor measurement prediction variance from the full communication, steady-state variance. For ease of notation, we introduce the transmit function

$$\gamma_j(k) := \mathbf{1}_{C_j(P(k)-\bar{P})C_j^{\mathrm{T}} \geq \delta_j}. \qquad (18)$$

Having established the transmit rule (17), we can now make the set $J(k)$ of all sensor measurements transmitted at time k precise:

$$J(k) = \{j \mid 1 \leq j \leq M,\, C_j\big(P(k)-\bar{P}\big)C_j^{T} \geq \delta_j\}. \qquad (19)$$

The matrices $\tilde{C}(k)$ and $\tilde{R}(k)$ for $k \in \mathbb{N}$ are well defined by (11), (19), and knowledge of $P(k) = \check{P}(k|k-1)$.

The index set $J(k)$ can have up to M elements; that is, more than one sensor may transmit their data at time k. In practice, this can for example be handled by low-level network protocols assigning priorities to the different sensors, as is the case in the application in [6]. We assume that the network bandwidth is sufficient for all selected sensors to transmit their data within on time interval. As we shall discuss in the next subsection, the transmit sequence (and hence the maximum number of sensors transmitting simultaneously) can be analyzed in advance.

The Kalman filter (12)–(16) together with the variance-based transmit decision (17) is referred to as the *event-based state estimator with variance-based triggering*. Since the Kalman filter (12)–(16) is the optimal Bayesian state estimator for any sequences $\tilde{C}(k)$ and $\tilde{R}(k)$, it is also optimal for those sequences given by (11) and (19). In other words, given the rule (17) (which captures the objective to use relevant measurements only), the filter (12)–(16) is the optimal state estimator for the estimation problem given by (7), (10), (11), and (19). Clearly, if $\delta_j = 0$ for all sensors, the full communication Kalman filter is recovered.

Remark 2: Alternative triggering rules than (17) may be useful for different scenarios. For example, when sensor noise is not stationary (i.e. $\mathrm{Var}\,[w_j(k)] = R_{jj}(k)$ instead of R_{jj}), the modified triggering rule

$$\text{transmit } y_j(k) \Leftrightarrow\ C_j P(k) C_j^{\mathrm{T}} + R_{jj}(k) \geq \delta_j \qquad (20)$$

can be useful for adapting the sensor transmit rates to varying $R_{jj}(k)$. For instance, if sensor j observes a change in sensor conditions leading to an updated $R_{jj}(k)$, it can take this into account in the transmit decision (20). The updated $R_{jj}(k)$ can be communicated alongside $y_j(k)$ at the next triggering instant, so that all estimators can use the updated variance from thereon for making predictions.

Remark 3: The triggering rule (17) uses constant rather than time-varying thresholds δ_j. This is partly motivated by practical considerations (constant thresholds are easier to implement), and also suggested by theoretical results on the optimal design of communication logics for related problems such as in [9], [19]. Therein, optimal event-triggering rules are designed such that a cost function, which captures estimation or control performance, is minimized. Typically, the optimal triggering rules are time-varying for finite horizon problems, while they are constant for infinite horizon problems, [9, p. 340]. Since we are mostly interested in long-term behavior (average communication rates and long-term estimation performance), and the system (7), (8) is time invariant, there is essentially no need to vary the threshold. For other applications, where conditions such as communication loads or plant parameters change, adaptive thresholds may be preferable.

C. Switching Riccati Equation

The update equation for the estimator prediction variance $P(k)$ is obtained by combining (11), (13), (14), (16), and (19):

$$\begin{aligned} P(k+1) = A\,P(k)\,A^{\mathrm{T}} + Q - A\,P(k)\,\check{C}^{\mathrm{T}}(P(k)) \\ \cdot \big(\check{C}(P(k))\,P(k)\,\check{C}^{\mathrm{T}}(P(k)) + \check{R}(P(k))\big)^{-1} \\ \cdot \check{C}(P(k))\,P(k)\,A^{\mathrm{T}} =: H(P(k)), \end{aligned} \qquad (21)$$

where $\check{C}(P(k)) := \tilde{C}(k)$ and $\check{R}(P(k)) := \tilde{R}(k)$ have been introduced to emphasize their dependence on $P(k)$ by (11) and (19); and $H(\cdot)$ denotes the map from $P(k)$ to $P(k+1)$. The system given by (7), (10), and (11) can be regarded as a switching system with modes given by the possible values of $J(k)$. The modes of the system are switched as a function of the prediction variance at the previous step through (19). Thus, (21) is a Riccati-type iteration with switching that depends on the variance at the previous step.

According to (21), the sequence $P(k)$ for $k \in \mathbb{N}^+$ can be computed from the problem data (A, C, Q, R, P_0), and the tuning parameters δ_j. Notice that this is fundamentally different from approaches such as [4], where the decision whether to transmit a measurement is based on the actual real-time measurement. Since the measurement is a random variable, the Kalman filter variables $P(k)$ and $\check{P}(k|k)$ become random variables themselves; whereas here, they are deterministic and can be computed off-line from the problem data. This will allow the analysis of the switching Riccati equation (21) in the following sections. Of course, the off-line analysis is only possible if the problem data is known ahead of time (it is not, for example, in the scenario discussed in Remark 2).

III. Illustrative Examples

We provide two examples to illustrate the behavior of the switching Riccati equation (21) for the event-based state estimator. The solutions are asymptotically periodic in both examples. Periodic solutions of the Riccati iteration correspond to periodic transmit sequences, which gives rise to a time-triggered implementation of the event-based design with low

complexity (the sensor nodes do not need to run a copy of the estimator then). Matlab files to reproduce the simulation results of this section are provided as supplementary material with this article.

A. Scalar Problem

Consider the system (7), (8) with a single sensor ($M = 1$), a scalar process ($n = 1$), and the parameters (small letters are used to indicate scalar quantities):

Example 1: $a = 1.2$, $c = q = r = 1$, $\delta = 3$, $p_0 = \bar{p}$.
Figure 2(a) shows the results from simulating (21).

As expected, the prediction variance $p(k)$ grows at times where no measurement is available. Once the threshold is exceeded, a measurement is transmitted ($\gamma(k) = 1$) and the estimator variance drops. The solution in Fig. 2(a) asymptotically converges to a periodic solution with period $N = 3$.

Figures 2(b) and 2(c) illustrate that, for different values of δ (all other parameters are the same as in Example 1), asymptotically periodic solutions with very different periods can be obtained. Notice that the period does not vary monotonically with δ.

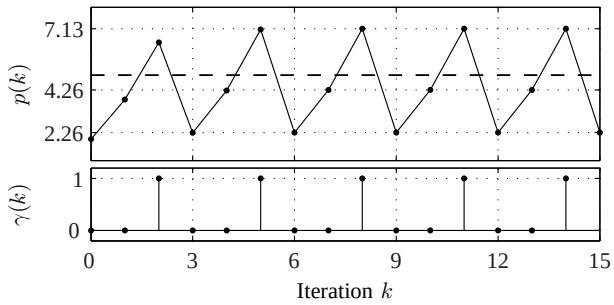
B. Multivariate Problem of the Balancing Cube

The event-based state estimation method was used in [6] to reduce the average communication in the networked control system of the Balancing Cube [7]. As an example for the multivariate version of (21), we recapitulate the simulation results which were used in [6] to find periodic transmit sequences. Please refer to [7] for a detailed description of the system and to [6] for details on the model and the design of the event-based state estimator.

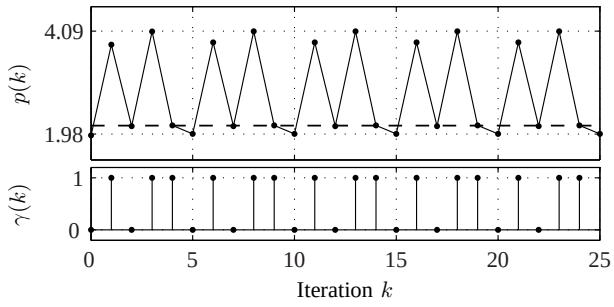
The model that is used for state estimation has $n = 8$ states. Noisy measurements of the states $x_1, \ldots, x_6$ and x_8 are available. The result of simulating the switching Riccati equation (21) for this system is shown in Fig. 3 (for iterations 1000 to 1150). The obtained solution converges to an N-cycle with period $N = 50$. Corresponding to this solution, the sensor measuring x_1 transmits every 50th time step (the same holds for the sensors measuring $x_2, \ldots, x_6$, whose variance is not shown in Fig. 3), and the measurement of x_8 is transmitted every 5th time step. This corresponds to a substantial reduction in average communication compared to the full communication Kalman filter. As is shown in [6], this reduction in communication comes at only a mild decrease in control performance when the estimate is used for feedback control.

IV. Asymptotic Periodicity for Scalar Problem

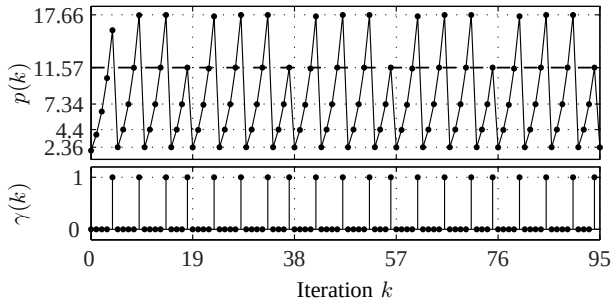
For the specific problems in the previous section, we observed in simulations that the Riccati iteration (21) approaches periodic solutions (i.e. convergence to periodic solutions was interpreted from simulation results rather than proven). In this section, we address the convergence problem and derive a theorem for the scalar version of (21) that guarantees asymptotic periodicity of the solution under certain assumptions to be derived in this section as well.



(a) $\delta = 3$



(b) $\delta = 0.2$



(c) $\delta = 9.6167$

Fig. 2. Simulation results for the scalar Example 1 and different values of the threshold parameter δ. The top graph of each sub-figure shows the variance iterates $p(k)$ (dots) and the transmit threshold $\bar{p}+\delta/c^2$ (dashed). The bottom graph shows the corresponding transmit sequence $\gamma(k)$. All solutions are asymptotically periodic with periods $N = 3, 5, 19$ from (a) to (c).

The question, under what conditions is the periodic transmission of sensor data the optimal solution for the event-based estimation problem posed by (7), (10), (11), and (19), is of theoretical interest for understanding the connection between time-triggered and event-based estimation. On the other hand, checkable conditions for asymptotic periodicity are also of practical value as they provide a means of identifying periodic solutions other than by simulating (21) and having to interpret the result (where it may happen that one has not simulated long enough to find a solution with a larger period).

The subject of study in this section is the scalar version of (21); that is, the nonlinear recursive equation

$$p(k+1) = a^2\, p(k) + q - \mathbf{1}_{c^2(p(k)-\bar{p})\geq\delta}\, \frac{a^2\, c^2\, p^2(k)}{c^2\, p(k) + r} \qquad (22)$$

$$p(0) = p_0 \geq 0, \qquad (23)$$

for the parameters $|a| > 1$, $c \neq 0$, $q > 0$, $r > 0$, $\delta > 0$ (small letters are used to indicate scalar quantities). We

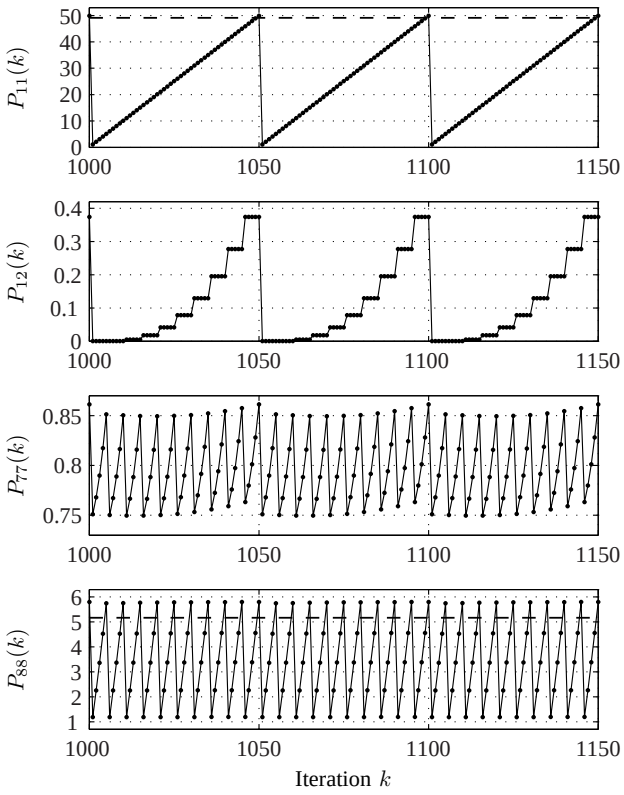


Fig. 3. Simulation results of the switching Riccati equation (21) for the Balancing Cube model after 1000 steps. Shown are some elements of $P(k)$ (dots) and the transmit threshold $\bar{P}_{jj} + \delta_j$ (dashed). Notice that there is no explicit threshold on $P_{12}(k)$ and $P_{77}(k)$. The solution is asymptotically periodic with period $N = 50$.

study the scalar problem (22) since it represents the simplest version of the matrix equation (21) that still exhibits its main characteristic, namely switching due to the variance-based trigger. Furthermore, we restrict attention to unstable dynamics ($|a| > 1$); in this case, communication of measurements is required for the estimation error variance to be bounded. We derive conditions that guarantee the solution of (22) to be asymptotically N-periodic, and give an algorithm to compute the period N. After some preliminaries in Sec. IV-A, we use an illustrative example in Sec. IV-B to outline the convergence proof, which then follows in Sec. IV-C to IV-F with the main result being stated in Sec. IV-F (Theorem 2).

A. Preliminaries

Since $q > 0$ and $r > 0$, (22) can equivalently be written as

$$\frac{p(k+1)}{q} = a^2 \frac{p(k)}{q} + 1 - \mathbf{1}_{\frac{p(k)}{q} - \frac{\bar{p}}{q} \geq \frac{\delta}{c^2 q}} \frac{a^2 \frac{c^2 q}{r} \left(\frac{p(k)}{q}\right)^2}{\frac{c^2 q}{r} \frac{p(k)}{q} + 1}. \quad (24)$$

By redefining $p(k)$, c^2, and δ as $p(k)/q$, $c^2 q/r$, and $\delta/(c^2 q)$, respectively, we can assume without loss of generality that $q = r = 1$. Henceforth, we study the iteration

$$p(k+1) = h(p(k)), \quad p(0) = p_0 \geq 0, \quad (25)$$

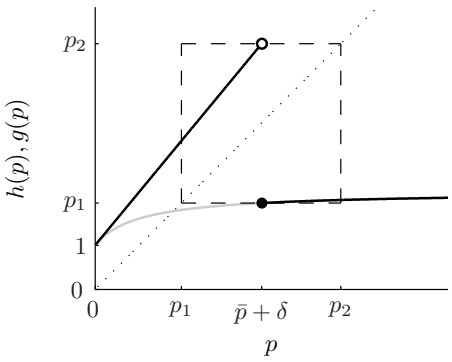


Fig. 4. The functions h (black) and g (gray). The filled circle indicates a closed interval boundary, whereas the unfilled circle indicates an open interval boundary. The dotted diagonal represents the identity map $p = p$. The intersection of g with the identity diagonal represents the solution $\bar{p}$ to the DARE (9). The dashed box represents the set $[p_1, p_2)$, which is invariant under h. For $p \geq \bar{p} + \delta$, $h(p) = g(p)$.

with h defined by

$$h : [0, \infty) \to [0, \infty)$$
$$p \mapsto a^2 p + 1 - \mathbf{1}_{p \geq \bar{p} + \delta} \frac{a^2 c^2 p^2}{c^2 p + 1} \quad (26)$$

with parameters $|a| > 1$, $c \neq 0$, $\delta > 0$; and with $\bar{p} = \text{DARE}(a, c, 1, 1)$. The graph of h is shown in Fig. 4 together with the graph of the function g,

$$g : [0, \infty) \to [0, \infty)$$
$$p \mapsto a^2 p + 1 - \frac{a^2 c^2 p^2}{c^2 p + 1}, \quad (27)$$

which represents the variance iteration of the full communication Kalman filter.

We summarize some properties of h, which will be useful later. The proof is straightforward and therefore omitted.

Proposition 1: Let $p_1 := h(\bar{p} + \delta)$ and $p_2 := a^2(\bar{p} + \delta) + 1$. The following properties of h hold:

(i) h is continuous and strictly increasing on $[p_1, \bar{p} + \delta)$ and on $[\bar{p} + \delta, p_2]$.
(ii) h is differentiable on $(p_1, \bar{p} + \delta)$ and on $(\bar{p} + \delta, p_2)$.
(iii) h is injective on $[p_1, p_2)$.
(iv) $h([p_1, p_2)) = [p_1, h(p_2)) \cup [h(p_1), p_2) \subseteq [p_1, p_2)$.
(v) $\forall p \in [0, \infty), \exists m \in \mathbb{N} : h^m(p) \in [p_1, p_2)$.

For $h([p_1, p_2)) \subseteq [p_1, p_2)$ in (iv), we also say that $[p_1, p_2)$ is an invariant set under h. From (iv) and (v), ultimate boundedness of the solution to (25) (in the sense of [30]) is immediate:

Corollary 1: For any initial condition $p_0 \geq 0$, the solution to (25) is ultimately bounded from above and below; that is, there exists $m \in \mathbb{N}$ such that $p_1 \leq p(k) < p_2$ for all $k \geq m$, with p_1 and p_2 as defined in Proposition 1.

Corollary 1 shows the effect of the threshold parameter δ on the estimation quality; in particular, the upper bound $p_2 := a^2(\bar{p} + \delta) + 1$ can directly be adjusted via this parameter. Furthermore, the corollary allows us to restrict attention to the interval $[p_1, p_2)$ for studying conditions for asymptotic periodicity in the following.

From (iii), the inverse of h exists on the range of h on $[p_1, p_2)$, which is $h([p_1, p_2)) = [p_1, h(p_2)) \cup [h(p_1), p_2)$ by

(iv). Hence, we define the inverse h^{-1} as

$$\begin{aligned} h^{-1} \; &: [p_1, h(p_2)) \cup [h(p_1), p_2) \to [p_1, p_2) \\ &\quad y \;\mapsto\; h^{-1}(y) \text{ such that } h(h^{-1}(y)) = y. \end{aligned} \tag{28}$$

The convergence proof in the following subsections is based on the *contraction mapping theorem*:

Theorem 1 (Contraction Mapping Theorem, [31]): Let $\|\cdot\|$ be a norm for $\mathbb{R}^n$ and S a closed subset of $\mathbb{R}^n$. Assume $f : S \to S$ is a contraction mapping: there is an L, $0 \le L < 1$, such that $\|f(p) - f(\tilde{p})\| \le L\|p - \tilde{p}\|$ for all $p, \tilde{p}$ in S. Then f has a unique fixed point p^* in S. Furthermore, if $p(0) \in S$ and we set $p(k+1) = f(p(k))$, then

$$\|p(k) - p^*\| \le \frac{L^k}{1-L}\|p(1) - p(0)\| \quad (k \ge 0). \tag{29}$$

Equation (29) implies that $p(k)$ converges to p^* as $k \to \infty$ for any $p(0) \in S$.

B. Illustrative Example and Outline of the Proof

We now illustrate, by means of Example 1, the main ideas that are used in Sec. IV-C to IV-F to prove asymptotic periodicity of (25).

The graph of h for the parameters of Example 1 is shown in Fig. 5. Since there is no intersection with the identity diagonal, h has no fixed point, as expected. The graph of h^3, which is depicted in Fig. 6, does, however, have three intersections in $[p_1, p_2)$ with the identity diagonal. Hence, h^3 has three fixed points in this interval corresponding to the 3-cycle shown in Fig. 2(a).

We illustrate below how Theorem 1 can be used to systematically prove that h^3 has these three fixed points, and that they are (locally) attractive. This approach is then generalized in Sec. IV-C to IV-F to general solutions of (25). To be able to apply Theorem 1 (with $n = 1$, $f = h^3$, and $\|\cdot\| = |\cdot|$), there are two key requirements:

(R1) a suitable closed set S that is invariant under h^3 must be constructed, and

(R2) h^3 must be a contraction mapping on S.

As it shall be seen later, the discontinuities of the function h^3 play a crucial role in the development. The function h^3 has two discontinuities, which can be seen as follows:

- $h(p)$ is continuous for all $p \in [p_1, p_2)$ except at $d_1 := \bar{p} + \delta$.
- $h^2(p) = h(h(p))$ is continuous at p if h is continuous at p and if h is continuous at $h(p)$, [32]. Hence, points of discontinuity are d_1 (discontinuity of h); and $d_2 \in [p_1, p_2)$ such that $\bar{p} + \delta = d_1 = h(d_2)$. Since $d_1 \in \text{dom}(h^{-1})$, the inverse h^{-1} exists and $d_2 = h^{-1}(d_1)$.
- Similarly, $h^3(p) = h(h^2(p))$ is continuous at p if h^2 is continuous at p and if h is continuous at $h^2(p)$. Points of discontinuity are d_1 and d_2 (discontinuities of h^2); and (potentially) $d_3 \in [p_1, p_2)$ such that $\bar{p} + \delta = d_1 = h^2(d_3) \Leftrightarrow d_2 = h(d_3)$. But since $d_2 \notin \text{dom}(h^{-1})$ (cf. Fig. 5), such a d_3 does not exist. Hence, h^3 has the discontinuities d_1 and d_2.

The discontinuities d_1 and d_2 subdivide $[p_1, p_2)$ into three disjoint subintervals: $[p_1, p_2) = I_3 \cup I_2 \cup I_1$ with $I_3 := [p_1, d_2)$, $I_2 := [d_2, d_1)$, and $I_1 := [d_1, p_2)$. Figure 7 illustrates where one of the subintervals, I_1, is mapped by repeated application of h. It can be seen that $h^3(I_1) \subseteq [d_1, p_2) = I_1$. Furthermore, since $h(p_2) < d_2$ (cf. Fig. 5), the same property holds for the closure of I_1; that is, $[d_1, p_2]$ is invariant under h^3,

$$h^3([d_1, p_2]) \subseteq [d_1, p_2]. \tag{30}$$

Notice that $h([d_1, p_2])$ and $h^2([d_1, p_2])$ (the same intervals as $h(I_1)$ and $h^2(I_1)$ in Fig. 7, but with closed right bounds) are closed intervals contained in I_3 and I_2, respectively. It can be shown that, under h^3, they are invariant and attractive for any point in I_3 and I_2, respectively. Hence, we can construct closed sets invariant under h^3 (requirement (R1)).

For (R2), we focus again on the interval I_1. Consider the derivative of h^3 on (d_1, p_2). By the chain rule, for $p \in (d_1, p_2)$,

$$\frac{d(h^3)}{dp}(p) = h'(h^2(p)) \cdot h'(h(p)) \cdot h'(p), \tag{31}$$

where h' refers to the first derivative $\frac{dh}{dp}$. Similar to the argumentation in Fig. 7, one can see that $h((d_1, p_2)) \subseteq (p_1, d_2)$ and $h^2((d_1, p_2)) \subseteq h((p_1, d_2)) \subseteq (d_2, d_1)$. From Fig. 5, it can be seen that $h'(p) = a^2$ for all $p \in (p_1, d_2) \cup (d_2, d_1)$ and that $h'(p) < g'(d_1)$ for all $p \in (d_1, p_2)$. Therefore, for $p \in (d_1, p_2)$, we get from (31) the following:

$$\frac{d(h^3)}{dp}(p) < a^2 \cdot a^2 \cdot g'(d_1) = a^4 g'(\bar{p} + \delta) = 0.084. \tag{32}$$

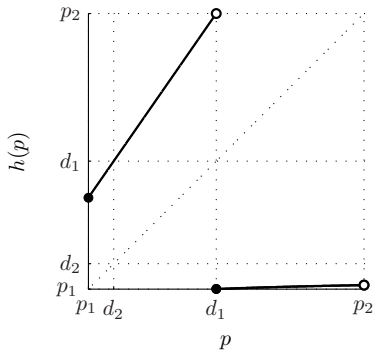


Fig. 5. The function h for $a = 1.2$, $c = 1$, $\delta = 3$ on the interval $[p_1, p_2) = [2.20, 8.13)$. The function has a discontinuity at $d_1 = \bar{p} + \delta = 4.95$. The slope of h is a^2 on (p_1, d_1) and bounded by $g'(d_1)$ on (d_1, p_2) (cf. Fig. 4).

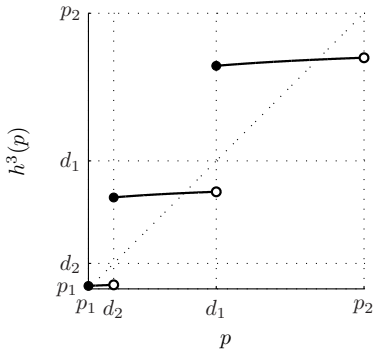


Fig. 6. The function h^3 for $a = 1.2$, $c = 1$, $\delta = 3$ on the interval $[p_1, p_2) = [2.20, 8.13)$. The function has two discontinuities at $d_1 = \bar{p} + \delta = 4.95$ and $d_2 = 2.74$.

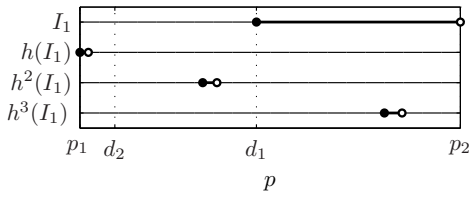


Fig. 7. Mapping of the interval I_1 under repeated application of h. On the top line, the interval $I_1 = [d_1, p_2)$ is shown as a thick line. This interval is mapped to $h(I_1) = [h(d_1), h(p_2)) = [p_1, h(p_2))$ (cf. Fig. 5), shown on the second line from above. Notice that the obtained interval is significantly shorter due to the slope of h being significantly less than one on $[d_1, p_2)$ (cf. Fig. 5). The intervals $h^2(I_1) = h(h(I_1))$ and $h^3(I_1) = h(h^2(I_1))$ (third and fourth line from above) are obtained accordingly. The interval length increases for the latter two mappings, since the slope of h is greater than one on $[p_1, d_1)$. Still, after one cycle of three mappings, the resulting interval is contained in the original one, i.e. $h^3(I_1) \subseteq I_1$.

From this, it follows (by the application of the mean value theorem, [32]) that for any closed interval $S \subseteq (d_1, p_2)$, the contraction mapping property in Theorem 1 holds with $L = a^4 g'(\bar{p}+\delta) < 1$. Even though the closed interval $[d_1, p_2]$ is not contained in (d_1, p_2), $\tilde{I}_1 := h^3([d_1, p_2])$ *is* contained (see Fig. 7). Furthermore, $\tilde{I}_1$ is itself invariant under h^3, which follows directly from (30): $h^3(\tilde{I}_1) = h^3(h^3([d_1, p_2])) \subseteq h^3([d_1, p_2]) = \tilde{I}_1$. Theorem 1 thus ensures that there exists a unique fixed point in $\tilde{I}_1$, and that every starting point in $\tilde{I}_1$ converges to this fixed point. Furthermore, since

$$h^3(I_1) = h^3([d_1, p_2)) \subseteq h^3([d_1, p_2]) = \tilde{I}_1, \qquad (33)$$

the fixed point is attractive for all points in the original interval I_1.

For the intervals I_2 and I_3, one can proceed similarly and, hence, show that every point in $[p_1, p_2)$ converges to a fixed point of h^3. Furthermore, we know by Corollary 1 that every solution to (25) ends up in $[p_1, p_2)$. Therefore, the solution to (25) for the considered example is asymptotically 3-periodic for any initial value $p_0 \geq 0$.

To treat the general case in the remainder of this section, we proceed analogously to this example. The construction of N closed subintervals of $[p_1, p_2)$ that are invariant under h^N proceeds in two steps. First, half-closed intervals I_i are generated that cover $[p_1, p_2)$ and possess the sought invariance property (Sec. IV-C). Second, closed intervals $\tilde{I}_i \subseteq I_i$ are constructed that inherit the invariance property from their supersets (Sec. IV-D). In Sec. IV-E, we show that h^N is a contraction mapping on these intervals, which then allows us (in Sec. IV-F) to apply Theorem 1 and conclude that solutions to (25) are asymptotically N-periodic.

C. Invariant Subintervals (Left-Closed, Right-Open)

Motivated by the example of the previous subsection, the left-closed, right-open intervals I_i are obtained by splitting up $[p_1, p_2)$ through a sequence of points $\{d_1, d_2, \dots\}$, $d_i \in [p_1, p_2)$, which represent discontinuities of h^N and are obtained by iteratively applying h^{-1} until some $d_i \notin \mathrm{dom}(h^{-1})$. We give an algorithm to compute these discontinuities:

Algorithm 1:

$d_1 := \bar{p} + \delta$
while $d_i \in \mathrm{dom}(h^{-1})$
  $d_{i+1} := h^{-1}(d_i)$
  increment i
end while
$N := i + 1$.

If there exists an $m \in \mathbb{N}$ such that $d_m \notin \mathrm{dom}(h^{-1})$, Algorithm 1 terminates, and the obtained sequence $\{d_1, d_2, \dots\}$ is finite. For all problems of an exhaustive search that we have conducted, this has actually been the case. For the purpose of this article, we assume henceforth that the algorithm terminates.

Assumption 1: Algorithm 1 terminates.

The assumption is essentially checked by running Algorithm 1 for a concrete problem; if the algorithm terminates, the assumption is true.

Proposition 2: Let $\mathcal{D}_i := \{d_1, \dots, d_i\}$ with d_i defined by Algorithm 1. The following statements hold:

(i) $\forall d_i, d_j \in \mathcal{D}_{N-1}$ with $i \neq j$, $d_i \neq d_j$.
(ii) $d_i \notin [h(p_2), h(p_1)]$, $\forall i < N-1$, $d_{N-1} \in [h(p_2), h(p_1))$.
(iii) h^i is continuous on $[p_1, p_2) \setminus \mathcal{D}_i$, $\forall i \leq N-1$, h^N is continuous on $[p_1, p_2) \setminus \mathcal{D}_{N-1}$.

Proof: (i): Proof by contradiction. Assume there exist $d_i, d_j \in \mathcal{D}_{N-1}$ with $i \neq j$ and $d_i = d_j$. Assume, without loss of generality, $j > i$ and let $m := j - i \leq N-2$. Then, from Algorithm 1, $d_i = d_j = h^{-1}(d_{j-1}) = h^{-2}(d_{j-2}) = \dots = h^{-m}(d_i)$; that is, the sequence of d_i's is periodic with period m, and Algorithm 1 never terminates, which contradicts with Assumption 1.

(ii): By Assumption 1, the sequence $\{d_1, d_2, \dots\}$ defined by Algorithm 1 is finite and equal to $\mathcal{D}_{N-1}$. Therefore, $d_i \in \mathrm{dom}(h^{-1})$ for all $i < N-1$ and $d_{N-1} \notin \mathrm{dom}(h^{-1})$. From $\mathrm{dom}(h^{-1}) = [p_1, h(p_2)) \cup [h(p_1), p_2)$ (see (28)), it follows that $d_i \notin [h(p_2), h(p_1))$ for all $i < N-1$. Furthermore, $d_i \neq h(p_1)$ can be seen by contradiction: assuming $d_i = h(p_1)$, it follows from $h(p_1) \in \mathrm{dom}(h^{-1})$ and $p_1 \in \mathrm{dom}(h^{-1})$ that there is $d_{i+2} \in \mathcal{D}_{N-1}$ with $d_{i+2} = h^{-2}(d_i) = h^{-1}(p_1) = \bar{p} + \delta = d_1$, which contradicts with (i).

Since h^{-1} maps to $[p_1, p_2)$ (see (28)), we have $d_{N-1} = h^{-1}(d_{N-2}) \in [p_1, p_2)$. Together with $d_{N-1} \notin \mathrm{dom}(h^{-1})$, this implies that $d_{N-1} \in [p_1, p_2) \setminus \big([p_1, h(p_2)) \cup [h(p_1), p_2)\big) = [h(p_2), h(p_1))$.

(iii): First, we prove by induction that h^i is continuous on $[p_1, p_2) \setminus \mathcal{D}_i$ for all $i \leq N-1$. From Proposition 1, (i), it follows that the statement is true for $i = 1$. Assume the statement holds for some $i \leq N-2$ (induction assumption (IA)); and consider

$$h^{i+1}(p) = h(h^i(p)), \quad p \in [p_1, p_2). \qquad (34)$$

If h^i is continuous at p and h is continuous at $h^i(p)$, then the composition h^{i+1} is continuous at p, [32]. Hence, h^{i+1} is guaranteed to be continuous on $[p_1, p_2)$ except for the points $\mathcal{D}_i$ and the point $\tilde{p}$ with $h^i(\tilde{p}) = d_1$ (d_1 is the discontinuity of h). But $h^i(\tilde{p}) = d_1 \Leftrightarrow \tilde{p} = h^{-i}(d_1) = d_{i+1}$ (since $i \leq N-2$, the i-times application of the inverse map, h^{-i}, is defined). Therefore, h^{i+1} is continuous on $[p_1, p_2) \setminus (\mathcal{D}_i \cup \{d_{i+1}\}) = [p_1, p_2) \setminus \mathcal{D}_{i+1}$.

Next, we prove that h^N is continuous on $[p_1, p_2) \setminus \mathcal{D}_{N-1}$. For this, consider

$$h^N(p) = h(h^{N-1}(p)), \quad p \in [p_1, p_2). \tag{35}$$

By the same argument as above, h^N is guaranteed to be continuous on $[p_1, p_2)$ except for the points $\mathcal{D}_{N-1}$ and the point $\tilde{p}$ with $h^{N-1}(\tilde{p}) = d_1 \Leftrightarrow h(\tilde{p}) = h^{-(N-2)}(d_1) = d_{N-1}$. But a point $\tilde{p}$ with $h(\tilde{p}) = d_{N-1}$ does not exist in $[p_1, p_2)$ since $d_{N-1} \in [h(p_2), h(p_1))$ (by (ii)), which is not in the range of h (by Proposition 1, (iv)). Therefore, h^N is continuous on $[p_1, p_2) \setminus \mathcal{D}_{N-1}$. ∎

The points $\mathcal{D}_{N-1}$ divide the interval $[p_1, p_2)$ into N subintervals $\mathcal{I} := \{I_1, \ldots, I_N\}$. The intervals are named such that I_i has d_i as a lower bound for $i \leq N-1$, and I_N has the lower bound p_1 (cf. Sec. IV-B). A formal definition of the intervals is given next. Let $\Pi : \{1, \ldots, N-1\} \to \{1, \ldots, N-1\}$ be a permutation of the d_i's such that

$$d_{\Pi(i)} < d_{\Pi(i+1)}, \quad \forall i \in \{1, \ldots, N-2\}. \tag{36}$$

Furthermore, let $\underline{i}$ and $\bar{i}$ be the indices of the smallest and greatest d_i, i.e. $\Pi(1) = \underline{i}$ and $\Pi(N-1) = \bar{i}$. Then define

$$I_i := [d_i, d_{\Pi(\Pi^{-1}(i)+1)}) \quad \forall i \leq N-1, i \neq \bar{i} \tag{37}$$

$$I_{\bar{i}} := [d_{\bar{i}}, p_2) \tag{38}$$

$$I_N := [p_1, d_{\underline{i}}); \tag{39}$$

that is, interval I_i has d_i as a lower bound (closed) and the next bigger element from $\mathcal{D}_{N-1}$ as an upper bound (open), except for the intervals at the boundaries of $[p_1, p_2)$. Since each interval is uniquely specified from (37)–(39) by either its lower or its upper bound, we sometimes omit either one of them and write $[d, *)$ or $[*, d)$. For the interior (the largest contained open interval) of I_i, we write $\text{int}(I_i)$.

Proposition 3: All intervals $I_i \in \mathcal{I}$ are mutually disjoint and nonempty.

Proof: Disjointness of the intervals is given by their construction and Proposition 2, (i). Furthermore, because of Proposition 2, (i), the intervals (37) are not empty. Since $d_{\bar{i}} \in \mathcal{D}_{N-1}$, it follows that $d_{\bar{i}} \in [p_1, p_2)$ and $d_{\bar{i}} < p_2$; therefore, interval $I_{\bar{i}}$ in (38) is not empty. To see that I_N in (39) is not empty, we assume that it is and lead this to a contradiction. From $[p_1, d_{\underline{i}}) = \emptyset$ it follows that $p_1 = d_{\underline{i}}$ ($p_1 > d_{\underline{i}}$ is not possible since $d_{\underline{i}} \in [p_1, p_2)$). From $d_{\underline{i}} = p_1 \in \text{dom}(h^{-1})$, it follows that $d_{\underline{i}+1}$ is defined by Algorithm 1 and $d_{\underline{i}+1} = h^{-1}(d_{\underline{i}}) = h^{-1}(p_1) = h^{-1}(h(\bar{p} + \delta)) = \bar{p} + \delta = d_1$. But $d_{\underline{i}+1} = d_1$ (with $\underline{i} \geq 1$) contradicts Proposition 2, (i). ∎

Proposition 4: The following statements hold:

(i) $h(I_N) \subseteq I_{N-1}$, $h(I_{N-1}) \subseteq I_{N-2}$, …, $h(I_2) \subseteq I_1$, and $h(I_1) \subseteq I_N$.

(ii) $h(\text{int}(I_N)) \subseteq \text{int}(I_{N-1})$, $h(\text{int}(I_{N-1})) \subseteq \text{int}(I_{N-2})$, …, $h(\text{int}(I_2)) \subseteq \text{int}(I_1)$, and $h(\text{int}(I_1)) \subseteq \text{int}(I_N)$.

The following lemma is used in the proof of this proposition (statements (i) and (ii)) and later in Sec. IV-D ((iii) and (iv)).

Lemma 1: Consider the collection $\mathcal{I} = \{I_1, \ldots, I_N\}$ of intervals I_i defined by (37)–(39); and let $\mathcal{I}_{\text{int}} := \{\text{int}(I_1), \ldots, \text{int}(I_N)\}$. The following statements hold:

(i) $\mathcal{I} \xrightarrow{h} \mathcal{I}$.

(ii) $\mathcal{I}_{\text{int}} \xrightarrow{h} \mathcal{I}_{\text{int}}$.

(iii) $I_{\bar{i}-N1} = \begin{cases} [d_{\bar{i}-1}, d_{N-1}) & \bar{i} > 1 \\ [p_1, d_{N-1}) & \bar{i} = 1. \end{cases}$

(iv) $\text{int}(I_{N-1}) = \begin{cases} (d_{N-1}, d_{\underline{i}-1}) & \underline{i} > 1 \\ (d_{N-1}, p_2) & \underline{i} = 1. \end{cases}$

Proof: The proof is given in Appendix A. ∎

Proof (Proposition 4): We present the proof of (i); the proof of (ii) is analogous and therefore omitted. From Lemma 1, (i), we know that, for any $I \in \mathcal{I}$, $h(I)$ is contained in an interval of $\mathcal{I}$. Since the intervals are disjoint (Proposition 3), there is exactly one interval that contains $h(I)$. Therefore, it suffices to consider only the lower bound of an interval to identify where the interval is mapped to.

Notice that, since h is strictly increasing (Proposition 1, (i)), for all $[a, b) \in \mathcal{I}$, $h([a, b)) = [h(a), \lim_{p \nearrow b} h(p))$. From Algorithm 1, it follows that $h(d_i) = d_{i-1}$ for all $i \in \{2, \ldots, N-1\}$. Since there is exactly one interval in $\mathcal{I}$ with d_{i-1} as lower bound, for all $i \in \{2, \ldots, N-1\}$,

$$h(I_i) = h([d_i, *)) = [d_{i-1}, *) \subseteq I_{i-1} \tag{40}$$

Similarly, since $h(d_1) = h(\bar{p} + \delta) = p_1$ by the definitions of d_1 and p_1, it follows that

$$h(I_1) = h([d_1, *)) = [p_1, *) \subseteq I_N \tag{41}$$

From Proposition 2, (ii), it follows that $h(p_1) \in [d_{N-1}, *) = I_{N-1}$. Therefore,

$$h(I_N) = h([p_1, *)) = [h(p_1), *) \subseteq I_{N-1}$$ ∎

The corollary follows directly from Proposition 4:

Corollary 2: The following statements hold:

(i) $h^N(I_i) \subseteq I_i \ \ \forall I_i \in \mathcal{I}$.

(ii) $h^N(\text{int}(I_i)) \subseteq \text{int}(I_i) \ \ \forall I_i \in \mathcal{I}$.

D. Invariant Closed Subintervals

The intervals $\mathcal{I}$ cover the whole domain of interest $[p_1, p_2)$, and they are invariant under h^N. However, closed intervals are required if Theorem 1 is to be applied. The proposition below states that such subintervals $\tilde{I}_i \subseteq I_i$ exist. A technical assumption is required for this proposition:

Assumption 2: $h(p_2) \neq d_{N-1}$.

Notice that with $h(p_2) \leq d_{N-1}$ by Proposition 2, (ii), this implies

$$h(p_2) < d_{N-1}. \tag{42}$$

Proposition 5: There exists a collection of intervals $\tilde{\mathcal{I}} = \{\tilde{I}_1, \tilde{I}_2, \ldots, \tilde{I}_N\}$ such that for all $i \in \{1, \ldots, N\}$ the following statements hold:

(i) $\tilde{I}_i$ is closed.

(ii) $\tilde{I}_i \subseteq \text{int}(I_i) \subseteq I_i$.

(iii) $h^N(\tilde{I}_i) \subseteq \tilde{I}_i$.

(iv) $h^{2N}(I_i) \subseteq \tilde{I}_i$.

Proof: The closed intervals $\tilde{I}_i$ can be constructed along the same lines as illustrated in Sec. IV-B, by taking the closure of the right-most interval $[d_{\bar{i}}, p_2]$ and considering its mappings $h^i([d_{\bar{i}}, p_2])$, $i \geq 1$, through h. The details of the proof are given in Appendix B. ∎

E. Contraction Mapping

In this section, we show that h^N is a contraction mapping. To this end, we first derive an upper bound less than one on the derivative of h^N for the interior of the intervals $\mathcal{I}$.

Proposition 6: h^N is differentiable on all open intervals $\text{int}(I_i)$, $I_i \in \mathcal{I}$. Furthermore, there exists an L, $0 \leq L < 1$, such that

$$\left|\frac{d(h^N)}{dp}(p)\right| < L \quad \forall p \in \text{int}(I_i), \forall I_i \in \mathcal{I}.$$

The following Lemma is needed in the proof.

Lemma 2: For all $p \in [p_1, \bar{p}+\delta)$, there exists an $m(p) \in \mathbb{N}^+$ such that

$$p, h(p), \ldots, h^{m(p)-1}(p) < \bar{p}+\delta \quad \text{and} \quad h^{m(p)}(p) \geq \bar{p}+\delta. \tag{43}$$

Furthermore, there exists an $\bar{N} \in \mathbb{N}^+$ (independent of p) such that $m(p) \leq \bar{N}$, and

$$a^{2\bar{N}} < a^2 \frac{\bar{p}+\delta}{p_1}. \tag{44}$$

Proof: The proof is given in Appendix C. ∎

The lemma says that if $p(0)$ starts anywhere in $[p_1, \bar{p}+\delta)$, there is a maximum number $\bar{N}$ of iterations (25), for which $p(k)$ remains in $[p_1, \bar{p}+\delta)$.

Proof (Proposition 6): Take any $I_i \in \mathcal{I}$ and $\tilde{p} \in \text{int}(I_i)$.

Differentiability: By Proposition 1, (ii), h is differentiable at $\tilde{p}$. Using the chain rule [32] and Proposition 4, (ii), it can be shown by induction that h^j is differentiable at $\tilde{p}$ for all $1 \leq j \leq N$.

Contraction mapping: By the chain rule,

$$\begin{aligned}(h^N)'(\tilde{p}) &= h'(h^{N-1}(\tilde{p})) \cdot (h^{N-1})'(\tilde{p}) \\ &= \prod_{j\in\{0,\ldots,N-1\}} h'(h^j(\tilde{p})) = \prod_{p\in\mathcal{P}} h'(p),\end{aligned} \tag{45}$$

with $\mathcal{P} := \{\tilde{p}, h(\tilde{p}), \ldots, h^{N-1}(\tilde{p})\}$. Notice from Proposition 4, (ii), that, for every point $p \in \mathcal{P}$, there is exactly one interval $I \in \mathcal{I}$ such that $p \in \text{int}(I)$.

Let $\mathcal{I}_L \subset \mathcal{I}$ denote the set of all intervals $I \in \mathcal{I}$ with $I < \bar{p}+\delta$ (intervals left of the discontinuity $\bar{p}+\delta$), and let $\mathcal{I}_R \subset \mathcal{I}$ denote the set of all $I \in \mathcal{I}$ with $I \geq \bar{p}+\delta$ (intervals right of the discontinuity $\bar{p}+\delta$). Furthermore, let N_L and N_R denote the number of elements in $\mathcal{I}_L$ and $\mathcal{I}_R$, respectively. Notice that $N_L \geq 1$ and $N_R \geq 1$. Then,

$$h'(p) = a^2 > 0 \qquad \forall p \in \text{int}(I),\ I \in \mathcal{I}_L, \tag{46}$$

follows directly from (26); and

$$0 < h'(p) = g'(p) < g'(\bar{p}+\delta) \qquad \forall p \in \text{int}(I),\ I \in \mathcal{I}_R, \tag{47}$$

where the first inequality follows from $g'(p) = \frac{a^2}{(c^2p+1)^2} > 0$, and the second inequality follows from g' being strictly decreasing, which is seen from $g''(p) = -\frac{2a^2c^2}{(c^2p+1)^3} < 0$. With these results, it follows from (45) that

$$0 < (h^N)'(\tilde{p}) < a^{2N_L} \left(g'(\bar{p}+\delta)\right)^{N_R}. \tag{48}$$

Since $a^2 > 1$ and $g'(\bar{p}+\delta) < 1$, whether the map h^N is contractive depends on the ratio of N_R to N_L, which is investigated next.

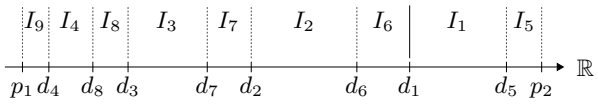


Fig. 8. Illustration of the intervals $\mathcal{I}$ obtained for the parameter values $a = 1.2$, $c = 1$, and $\delta = 9.6$ (for better visibility the relative scaling of the intervals has been adapted). There are two distinct interval subsequences satisfying (49): $\underline{\mathcal{I}}_1 = \{I_4, I_3, I_2\}$ and $\underline{\mathcal{I}}_2 = \{I_9, I_8, I_7, I_6\}$.

Define a subset $\underline{\mathcal{I}} \subset \mathcal{I}$ as a maximum sequence of m intervals $I_\ell, I_{\ell-N1}, \ldots$ all being left of $\bar{p}+\delta$:

$$\begin{aligned}\underline{\mathcal{I}} := &\big\{I_\ell, I_{\ell-N1}, \ldots, I_{\ell-N(m-1)}\big\},\ m \leq N_L, \\ &\text{such that } I_\ell, I_{\ell-N1}, \ldots, I_{\ell-N(m-1)} \in \mathcal{I}_L, \\ &\text{and } I_{\ell-N(N-1)}, I_{\ell-Nm} \in \mathcal{I}_R,\end{aligned} \tag{49}$$

Let there be $\kappa \geq 1$ distinct interval sequences (49), which we call $\underline{\mathcal{I}}_1, \ldots, \underline{\mathcal{I}}_\kappa$ with $m_1, \ldots, m_\kappa$ their numbers of elements, respectively. Notice that $N_L = m_1 + \cdots + m_\kappa$. An example with two interval sequences $\underline{\mathcal{I}}_1$, $\underline{\mathcal{I}}_2$ is provided in Fig. 8.

From Lemma 2, it follows that $m_j \leq \bar{N}$ for all $j \leq \kappa$ ($\bar{N}$ is as defined in Lemma 2), and

$$N_L = m_1 + \cdots + m_\kappa \leq \kappa\bar{N}. \tag{50}$$

For each sequence of intervals $\underline{\mathcal{I}}_j$, $j \leq \kappa$, there is at least one distinct interval $I \in \mathcal{I}_R$ (namely, $I_{\ell-Nm}$); hence,

$$N_R \geq \kappa. \tag{51}$$

Combining (50) and (51), we obtain the sought bound on the ratio of N_L and N_R: $N_L \leq \kappa\bar{N} \leq N_R\bar{N}$. With this result, we can rewrite (48),

$$\begin{aligned}0 < (h^N)'(\tilde{p}) &< a^{2N_L}\, a^{2(N_R\bar{N}-N_L)} \left(g'(\bar{p}+\delta)\right)^{N_R} \\ &= \left(a^{2\bar{N}}\, g'(\bar{p}+\delta)\right)^{N_R}.\end{aligned} \tag{52}$$

Using (44) from Lemma 2, we get

$$a^{2\bar{N}} g'(\bar{p}+\delta) < a^2 \frac{\bar{p}+\delta}{p_1}\, g'(\bar{p}+\delta). \tag{53}$$

The right-hand side in (53) depends on the problem parameters, and it can be shown to be less than one (the proof is omitted due to space constraints; the symbolic calculation is provided in a supplementary Matlab script). With this, the statement of Proposition 6 follows from (52) with $L := \left(a^{2\bar{N}}\, g'(\bar{p}+\delta)\right)^{N_R} < 1$. ∎

From Proposition 6 and the mean value theorem [32], it follows:

Corollary 3: h^N is a contraction mapping on any interval of $\tilde{\mathcal{I}}$ (defined in Proposition 5); that is, there exists an L, $0 \leq L < 1$, such that

$$|h^N(p) - h^N(\tilde{p})| \leq L|p - \tilde{p}| \quad \forall p, \tilde{p} \in \tilde{I}_i, \forall \tilde{I}_i \in \tilde{\mathcal{I}}.$$

F. Main Result

Equipped with the results of the previous subsections, we can now state the main result of this section:

Theorem 2: Under Assumptions 1 and 2, the solution to (25) is asymptotically N-periodic for any initial condition $p_0 \geq 0$.

Proof: By Corollary 1, it follows that there exists an $m_1 \in \mathbb{N}$ such that

$$h^{m_1 N}(p_0) \in [p_1, p_2). \tag{54}$$

Since the disjoint intervals $\mathcal{I}$ cover $[p_1, p_2)$, there exists a unique $i \in \{1, \ldots, N\}$ such that

$$h^{m_1 N}(p_0) \in I_i. \tag{55}$$

By Proposition 5, (iv),

$$h^{(m_1+2)N}(p_0) \in \tilde{I}_i. \tag{56}$$

From Proposition 5, (i) and (iii), Corollary 3, and Theorem 1, it follows that there exists a unique fixed point p_i^* of h^N (hence, an N-periodic point of (25)) in $\tilde{I}_i$ and that, for all $\tilde{p} \in \tilde{I}_i$,

$$\lim_{m\to\infty} h^{mN}(\tilde{p}) = p_i^*. \tag{57}$$

In particular, for $\tilde{p} = h^{(m_1+2)N}(p_0)$ and by (56),

$$\begin{aligned}\lim_{m\to\infty} h^{mN}\big(h^{(m_1+2)N}(p_0)\big) &= \lim_{m\to\infty} h^{(m_1+2+m)N}(p_0) \\ &= \lim_{m\to\infty} h^{mN}(p_0) = p_i^*.\end{aligned}$$

■

V. Discussion

The proposed method for event-based state estimation is a direct extension of the classic Kalman filter to a distributed estimation problem with costly communication. Starting from the design of a discrete-time Kalman filter with access to all sensor measurements at every sampling time, the presented method allows the designer to trade off communication requirements with estimation performance by selecting suitable thresholds for each sensor (these thresholds are the only additional tuning parameters).

The presented event-based estimator is the optimal Bayesian estimator given the triggering policy that a measurement is transmitted if, and only if, its prediction variance exceeds a given threshold (equation (17)). Thus, equation (17) represents a soft constraint on communication, in that transmission of a measurement only occurs if otherwise the measurement cannot be predicted well enough. A focus of this article is on the study of the asymptotic properties of the resulting variance recursion (the switching Riccati equation (21)), and on proving its convergence to periodic solutions for the special case of an unstable scalar process (Theorem 2). The event-based estimator variance being periodic in the limit means that a periodic, time-based transmit schedule is optimal in the above sense. The result thus establishes an interesting link between event-based and time-based optimal state estimation.

Assumption 1 is an important assumption in the derivation of Theorem 2. In all simulations that we performed, this assumption was satisfied. Based on this observation and preliminary analysis not presented herein, we conjecture that Assumption 1 holds true for almost all values of the problem parameters (in particular, the threshold δ). However, establishing this result is an open question beyond the scope of this article. As is, Theorem 2 provides a sufficiency test for periodicity (if the assumptions are satisfied, global convergence to a unique periodic solution with a known period is guaranteed) as an alternative to simulating the Riccati equation and having to interpret the result. The above conjecture has a practical ramification: if the algorithm does not terminate "quickly," one can perturb the parameter δ slightly (which affects estimation quality only insignificantly as by Corollary 1) and restart the algorithm.

While we propose variance-based triggering as a framework for reduced communication state estimation for systems with multiple sensors and states (and also applied it to such a system in [6]), the analysis herein focuses on scalar systems with a single sensor and, in particular, on the analysis of the corresponding scalar Riccati equation (22). Whether the results of asymptotic periodicity (Theorem 2) and also ultimate boundedness (Corollary 1) generalize to the matrix case (21) are open questions.

When a periodic solution to the switching Riccati equation is found (for both the scalar and the matrix case), sensor transmissions can be implemented as time-based, periodic schedules. The rate of the periodic transmission is, however, not a design parameter as in traditional time-sampled estimation, but obtained from an event-based approach, where the designer specifies tolerable bounds on the estimation variance. Hence, the method can be used as a tool for designing periodic sensor schedules for a multi-rate sensor system.

Alternatively, the estimation method can be implemented as shown in Fig. 1, where each sensor computes the variance of the estimator on-line and uses it to decide whether or not to broadcast its measurement. This way, sensor transmit rates can adapt in real time to unforeseen events (for example, varying sensor noise conditions as discussed in Remark 2, or packet drops). This implementation of estimation with variance-based triggering is in accordance with the common understanding of event-based methods (where events are generated on-line and can respond to unexpected changes), whereas with the periodic implementation mentioned above, transmission instants are computed off-line and can thus not adapt in real-time.

Various extensions of the presented event-based estimation approach are conceivable. Instead of making the transmit decision based on a measurement's prediction variance, it could be based on its prediction error; that is, the difference between the measurement and its prediction mean (see [4], [5]). A promising approach is to combine the two methods by augmenting fixed minimum sensor communication rates to keep the variance bounded with triggering thresholds on real-time prediction errors. If improving the estimator performance locally on each agent is of interest (for example, when the estimate is used in feedback for controlling an actuator), a second estimator can be used to compute an improved estimate from using the data received from the network and, additionally, exploiting its local sensor data at every time step (see [4]).

Appendix A
Proof of Lemma 1

For the proof of Lemma 1 at the end of this section, we need the following lemma and corollary.

Lemma 3: Let $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$ be a collection of nonempty, mutually disjoint intervals $I_i := [a_i, b_i)$ (or $I_i := (a_i, b_i)$) with $a_i, b_i \in \mathbb{R}$. A unique representation of $\mathcal{I}$ is given by the sets $\mathcal{L} = \{a_1, a_2, \ldots, a_N\}$ and $\mathcal{U} = \{b_1, b_2, \ldots, b_N\}$ of all lower and upper bounds, respectively, in the following sense: the collection $\bar{\mathcal{I}} := \{\bar{I}_1, \bar{I}_2, \ldots, \bar{I}_N\}$ of intervals constructed such that, for all i, j with $1 \le i, j \le N$,

$$\bar{I}_i := [\alpha_i, \beta_i) \text{ (or } \bar{I}_i := (\alpha_i, \beta_i)),\ \alpha_i \in \mathcal{L},\ \beta_i \in \mathcal{U}, \tag{58}$$

$$\bar{I}_i \neq \emptyset, \quad \text{and} \quad \bar{I}_i \cap \bar{I}_j = \emptyset, \tag{59}$$

exists, it is unique, and $\bar{\mathcal{I}} = \mathcal{I}$.

This lemma is useful, since it allows one to work with the (unordered) sets of interval bounds $\mathcal{L}$ and $\mathcal{U}$ instead of the actual intervals. The unique relationship between the bounds (which lower bound belongs to which upper bound) essentially follows from all intervals being disjoint and nonempty.

Proof: We present the proof simultaneously for the case of left-closed, right-open intervals $\bar{I}_i = [\alpha_i, \beta_i)$ and for the case of open intervals $\bar{I}_i = (\alpha_i, \beta_i)$ (where required to distinguish the two, the latter case is augmented in square brackets).

Since, for all $i \le N$, $I_i \in \mathcal{I}$ is nonempty, $a_i < b_i$. Since the intervals $\mathcal{I}$ are mutually disjoint, there exists a permutation of indices $\tilde{\Pi} : \{1, \ldots, N\} \to \{1, \ldots, N\}$ such that

$$a_{\tilde{\Pi}(1)} < b_{\tilde{\Pi}(1)} \le a_{\tilde{\Pi}(2)} < b_{\tilde{\Pi}(2)} \le \cdots \le a_{\tilde{\Pi}(N)} < b_{\tilde{\Pi}(N)}.$$

Assume w.l.o.g. (by renaming of the intervals in $\mathcal{I}$) that

$$a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_N < b_N. \tag{60}$$

Notice that the choice

$$\alpha_i = a_i \quad \text{and} \quad \beta_i = b_i \quad \text{for } 1 \le i \le N \tag{61}$$

satisfies (58), (59), and $\bar{\mathcal{I}} = \mathcal{I}$ trivially. Hence, a collection of intervals $\bar{\mathcal{I}}$ satisfying (58) and (59) exists; it remains to show that the choice (61) is unique.

We first show that, for any $a_i \in \mathcal{L}$, there is exactly one interval in $\bar{\mathcal{I}}$ that has a_i as a lower bound. We will show this by contradiction.

- Assume there is more than one interval with a_i as a lower bound; that is, there are $[a_i, b_j)$, $[a_i, b_\ell) \in \bar{\mathcal{I}}$ [$(a_i, b_j), (a_i, b_\ell) \in \bar{\mathcal{I}}$] with $b_j, b_\ell \in \mathcal{U}$ and $b_j > a_i$, $b_\ell > a_i$ (otherwise the intervals would be empty, which contradicts with (59)). But then, $[a_i, b_j) \cap [a_i, b_\ell) = [a_i, \min(b_j, b_\ell)) \neq \emptyset$ [$(a_i, b_j) \cap (a_i, b_\ell) = (a_i, \min(b_j, b_\ell)) \neq \emptyset$], which contradicts with (59).
- Assume there is no interval in $\bar{\mathcal{I}}$ that has a_i as a lower bound. Then, there can only be $N-1$ intervals in total, since it follows from the previous discussion that each of the remaining $a_j \in \mathcal{L} \setminus \{a_i\}$ can be chosen at most once as a lower bound. This contradicts with (58) (the collection $\bar{\mathcal{I}}$ having N elements).

Analogously, it can be shown that, for any $b_i \in \mathcal{U}$, there is exactly one interval in $\bar{\mathcal{I}}$ that has b_i as an upper bound. The detailed proof is omitted.

Now, take $\alpha_i = a_i$ for any $i \in \{1, \ldots, N\}$. From the discussion above, it follows that there is an interval $[\alpha_i, \beta_i) \in \bar{\mathcal{I}}$ [$(\alpha_i, \beta_i) \in \bar{\mathcal{I}}$], $\beta_i \in \mathcal{U}$. We prove by contradiction that this implies $\beta_i = b_i$, and, hence, that the choice (61) is unique.

Assume $\beta_i = b_j$, $b_j \neq b_i$. Then, from the above discussion, there exists also an interval $[a_\ell, b_i) \in \bar{\mathcal{I}}$ [$(a_\ell, b_i) \in \bar{\mathcal{I}}$], $a_\ell \in \mathcal{L}$. For $[a_i, b_j)$ [(a_i, b_j)] to be nonempty, it follows that $a_i < b_j$, which implies by (60) that $b_i \le b_j$. Similarly, for $[a_\ell, b_i)$ [(a_ℓ, b_i)] to be nonempty, $a_\ell < b_i$ implying $a_\ell \le a_i$. But then, $[a_i, b_j) \cap [a_\ell, b_i) = [a_i,\ b_i) \neq \emptyset$ [$(a_i, b_j) \cap (a_\ell, b_i) = (a_i,\ b_i) \neq \emptyset$], which contradicts (59). ∎

Corollary 4: Let $\mathcal{I}_1$, $\mathcal{I}_2$ be two collections of nonempty and mutually disjoint intervals. Let $\mathcal{L}_1$ and $\mathcal{U}_1$ be the sets of lower and upper bounds, respectively, of $\mathcal{I}_1$; and let $\mathcal{L}_2$ and $\mathcal{U}_2$ be the sets of lower and upper bounds, respectively, of $\mathcal{I}_2$. If $\mathcal{L}_1 = \mathcal{L}_2$ and $\mathcal{U}_1 = \mathcal{U}_2$, then $\mathcal{I}_1 = \mathcal{I}_2$.

Proof: Follows directly from Lemma 3. ∎

Proof (Lemma 1): (i), (ii): Statements (i) and (ii) are treated simultaneously ((ii) in brackets where required).

By Proposition 3, the intervals $\mathcal{I} = \{I_1, I_2, \ldots, I_N\} = \{[p_1, d_{\Pi(1)}), [d_{\Pi(1)}, d_{\Pi(2)}), \ldots, [d_{\Pi(N-1)}, p_2)\}$ are mutually disjoint and nonempty. Therefore, also the intervals $\mathcal{I}_{\text{int}} = \{\text{int}(I_1), \text{int}(I_2), \ldots, \text{int}(I_N)\} = \{(p_1, d_{\Pi(1)}), (d_{\Pi(1)}, d_{\Pi(2)}), \ldots, (d_{\Pi(N-1)}, p_2)\}$ are mutually disjoint and nonempty. Hence, by Lemma 3, $\mathcal{I}$ [$\mathcal{I}_{\text{int}}$] is uniquely represented by the sets of lower and upper bounds

$$\mathcal{L} = \{p_1, d_{\Pi(1)}, \ldots, d_{\Pi(N-1)}\} = \{p_1, d_1, \ldots, d_{N-1}\}, \tag{62}$$

$$\mathcal{U} = \{d_{\Pi(1)}, \ldots, d_{\Pi(N-1)}, p_2\} = \{p_2, d_1, \ldots, d_{N-1}\} \tag{63}$$

and the conditions (58) and (59) (note that the sets $\mathcal{L}$ and $\mathcal{U}$ are the same for $\mathcal{I}$ and $\mathcal{I}_{\text{int}}$, but (58) is different).

Define the collection of images of h on $\mathcal{I}$ [$\mathcal{I}_{\text{int}}$] as $\mathcal{I}_h := \{h([p_1, d_{\Pi(1)})), h([d_{\Pi(1)}, d_{\Pi(2)})), \ldots, h([d_{\Pi(N-1)}, p_2))\}$ [$\mathcal{I}_{\text{int},h} = \{h((p_1, d_{\Pi(1)})), h((d_{\Pi(1)}, d_{\Pi(2)})), \ldots, h((d_{\Pi(N-1)}, p_2))\}$]. Hence, by definition,

$$\mathcal{I} \xrightarrow{h} \mathcal{I}_h \qquad [\mathcal{I}_{\text{int}} \xrightarrow{h} \mathcal{I}_{\text{int},h}]. \tag{64}$$

Since, by Proposition 1, (i), h is strictly increasing on each $I_i = [a, b) \in \mathcal{I}$ [$I_i = (a, b) \in \mathcal{I}_{\text{int}}$], it holds that $h([a, b)) = [h(a), \lim_{p \nearrow b} h(p))$ [$h((a, b)) = (h(a), \lim_{p \nearrow b} h(p))$]. Therefore, the sets of lower and upper bounds of $\mathcal{I}_h$ [$\mathcal{I}_{\text{int},h}$] are given by

$$\begin{aligned}\mathcal{L}_h &:= \{h(a) \mid a \in \mathcal{L}\} = \{h(p_1), h(d_1), h(d_2), \ldots, h(d_{N-1})\} \\ &= \{h(p_1), p_1, d_1, \ldots, d_{N-2}\},\end{aligned} \tag{65}$$

$$\begin{aligned}\mathcal{U}_h &:= \{\lim_{p \nearrow b} h(p) \mid b \in \mathcal{U}\} \\ &= \{h(p_2), \lim_{p \nearrow d_1} h(p), h(d_2), \ldots, h(d_{N-1})\} \\ &= \{h(p_2), p_2, d_1, \ldots, d_{N-2}\},\end{aligned} \tag{66}$$

where we used the facts that h is continuous from the right at all $a \in \mathcal{L}$ and continuous from the left at all $b \in \mathcal{U} \setminus \{d_1\}$; and that

$$h(d_1) = h(\bar{p} + \delta) = p_1 \qquad \text{(def. of } p_1\text{)}, \tag{67}$$

$$h(d_i) = d_{i-1},\ 2 \le i \le N-1 \qquad \text{(by Alg. 1)}, \tag{68}$$

$$\lim_{p \nearrow d_1} h(p) = a^2(\bar{p} + \delta) + 1 = p_2 \qquad \text{(def. of } p_2\text{)}. \tag{69}$$

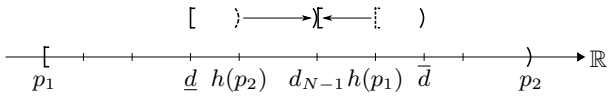


Fig. 9. Illustration of the enlargement of the intervals $[\underline{d}, h(p_2))$ and $[h(p_1), \overline{d})$ to $[\underline{d}, d_{N-1})$ and $[d_{N-1}, \overline{d})$. The points unspecified are elements from $\{d_1, \ldots, d_{N-2}\}$. All intervals remain nonempty and mutually disjoint.

Since h is injective (Proposition 1, (iii)), $h(I_1 \cap I_2) = h(I_1) \cap h(I_2)$ holds for any $I_1, I_2 \subseteq [p_1, p_2)$, [33]. From this, and the fact that the intervals $\mathcal{I}$ [$\mathcal{I}_{\text{int}}$] are disjoint, it follows that the mapped intervals $\mathcal{I}_h$ [$\mathcal{I}_{\text{int},h}$] are also disjoint. Furthermore, since h is not constant on any interval $I \in \mathcal{I}$ (it is strictly increasing by Proposition 1, (i)), the intervals $\mathcal{I}_h$ [$\mathcal{I}_{\text{int},h}$] are all nonempty. Hence, by Lemma 3, $\mathcal{I}_h$ [$\mathcal{I}_{\text{int},h}$] is uniquely represented by $\mathcal{L}_h$ and $\mathcal{U}_h$. Notice that $\mathcal{L}_h$ and $\mathcal{U}_h$ have the same elements as $\mathcal{L}$ and $\mathcal{U}$ except for $h(p_1)$ and $h(p_2)$ in $\mathcal{L}_h$ and $\mathcal{U}_h$, and d_{N-1} in $\mathcal{L}$ and $\mathcal{U}$. We show next that the intervals $\mathcal{I}_h$ [$\mathcal{I}_{\text{int},h}$] are contained in the intervals of $\mathcal{I}$ [$\mathcal{I}_{\text{int}}$].

To see this, notice first that the elements of $\mathcal{L}_h \cup \mathcal{U}_h \cup \mathcal{L} \cup \mathcal{U} = \{p_1, p_2, h(p_1), h(p_2), d_1, \ldots, d_{N-1}\}$ have the following order relation:

$$p_1 \leq \underbrace{\cdots\cdots}_{\text{other } d_i\text{'s}} < h(p_2) \leq d_{N-1} < h(p_1) < \underbrace{\cdots\cdots}_{\text{other } d_i\text{'s}} < p_2, \tag{70}$$

because

$$\begin{aligned}
&p_1 < h(p_2) && (\bar{p}+\delta < p_2 \text{ and Prop. 1(i)}),\\
&h(p_1) < p_2 && (\text{Prop. 1, (iv)}),\\
&h(p_2) \leq d_{N-1} < h(p_1) && (\text{Prop. 2, (ii)}),\\
&d_i \in [p_1, h(p_2)) \cup (h(p_1), p_2), && \forall i \in \{1, \ldots, N-2\}\\
& && (\text{Prop. 2, (ii)}).
\end{aligned}$$

Therefore, the upper bound of $[*, h(p_2)) \in \mathcal{I}_h$ [$(*, h(p_2)) \in \mathcal{I}_{\text{int},h}$] can be changed to d_{N-1}, and the lower bound of $[h(p_1), *) \in \mathcal{I}_h$ [$(h(p_1), *) \in \mathcal{I}_{\text{int},h}$] to d_{N-1}, without affecting the mutual disjointness and the nonemptiness of the intervals. This modification of the intervals is illustrated in Fig. 9, and we make it formal next.

Let $\underline{d}$ be the lower bound of $[*, h(p_2)) \in \mathcal{I}_h$ [$(*, h(p_2)) \in \mathcal{I}_{\text{int},h}$], and let $\overline{d}$ be the upper bound of $[h(p_1), *) \in \mathcal{I}_h$ [$(h(p_1), *) \in \mathcal{I}_{\text{int},h}$]. Then, define

$$\begin{aligned}
\tilde{\mathcal{I}}_h := &\big\{I \in \mathcal{I}_h \mid I \neq [\underline{d}, h(p_2)) \text{ and } I \neq [h(p_1), \overline{d})\big\}\\
&\cup \big\{[\underline{d}, d_{N-1}),\, [d_{N-1}, \overline{d})\big\},
\end{aligned} \tag{71}$$

that is, $\tilde{\mathcal{I}}_h$ has the same elements as $\mathcal{I}_h$ except for the replacements $[\underline{d}, h(p_2)) \to [\underline{d}, d_{N-1})$ and $[h(p_1), \overline{d}) \to [d_{N-1}, \overline{d})$. Similarly, define

$$\begin{aligned}
\tilde{\mathcal{I}}_{\text{int},h} := &\big\{I \in \mathcal{I}_{\text{int},h} \mid I \neq (\underline{d}, h(p_2)) \text{ and } I \neq (h(p_1), \overline{d})\big\}\\
&\cup \big\{(\underline{d}, d_{N-1}),\, (d_{N-1}, \overline{d})\big\}.
\end{aligned} \tag{72}$$

Since, from (70), $[\underline{d}, h(p_2)) \subseteq [\underline{d}, d_{N-1})$ [$(\underline{d}, h(p_2)) \subseteq (\underline{d}, d_{N-1})$] and $[h(p_1), \overline{d}) \subseteq [d_{N-1}, \overline{d})$ [$(h(p_1), \overline{d}) \subseteq (d_{N-1}, \overline{d})$], it follows from (64) that

$$\mathcal{I} \xrightarrow{h} \tilde{\mathcal{I}}_h \qquad [\mathcal{I}_{\text{int}} \xrightarrow{h} \tilde{\mathcal{I}}_{\text{int},h}]. \tag{73}$$

The lower and upper bounds of $\tilde{\mathcal{I}}_h$ ($\tilde{\mathcal{I}}_{\text{int},h}$) are given by

$$\tilde{\mathcal{L}}_h := \big\{d_{N-1}, p_1, d_1, \ldots, d_{N-2}\big\}, \tag{74}$$

$$\tilde{\mathcal{U}}_h := \big\{d_{N-1}, p_2, d_1, \ldots, d_{N-2}\big\}. \tag{75}$$

Since the intervals $\tilde{\mathcal{I}}_h$ [$\tilde{\mathcal{I}}_{\text{int},h}$] are nonempty and mutually disjoint, and $\tilde{\mathcal{L}}_h = \mathcal{L}$ and $\tilde{\mathcal{U}}_h = \mathcal{U}$, it follows from Corollary 4 that $\tilde{\mathcal{I}}_h = \mathcal{I}$ [$\tilde{\mathcal{I}}_{\text{int},h} = \mathcal{I}_{\text{int}}$]. Using this result, statements (i) and (ii) are given by (73).

(iii): First, notice that $1 \leq \bar{i} \leq N-1$ and

$$h(I_{\bar{i}}) \underset{(38)}{=} h([d_{\bar{i}}, p_2)) \underset{(67),(68)}{=} \begin{cases} [d_{\bar{i}-1}, h(p_2)) & \text{if } \bar{i} > 1\\ [p_1, h(p_2)) & \text{if } \bar{i} = 1. \end{cases} \tag{76}$$

Since $h(I_{\bar{i}}) \in \mathcal{I}_h$, it follows that

$$\underline{d} = \begin{cases} d_{\bar{i}-1} & \text{if } \bar{i} > 1\\ p_1 & \text{if } \bar{i} = 1 \end{cases} \tag{77}$$

($\underline{d}$ has been defined above as the lower bound of the (unique) interval in $\mathcal{I}_h$ that has $h(p_2)$ as an upper bound). Notice that in $\mathcal{I}$, there is exactly one interval with $d_{\bar{i}-1}$ as lower bound (for $\bar{i} > 1$) and exactly one interval with p_1 as lower bound. Therefore, it follows from $[\underline{d}, d_{N-1}) \in \tilde{\mathcal{I}}_h = \mathcal{I}$ (see (71)), and (37)–(39) that

$$\begin{aligned}
[\underline{d}, d_{N-1}) &= \begin{cases} [d_{\bar{i}-1}, d_{N-1}) & \text{if } \bar{i} > 1\\ [p_1, d_{N-1}) & \text{if } \bar{i} = 1 \end{cases} = \begin{cases} I_{\bar{i}-1} & \text{if } \bar{i} > 1\\ I_N & \text{if } \bar{i} = 1 \end{cases}\\
&= I_{\bar{i}-_N 1}.
\end{aligned} \tag{78}$$

(iv): Notice that $1 \leq \underline{i} \leq N-1$ and

$$h(\text{int}(I_N)) \underset{(39)}{=} h((p_1, d_{\underline{i}})) \underset{(68),(69)}{=} \begin{cases} (h(p_1), d_{\underline{i}-1}) & \text{if } \underline{i} > 1\\ (h(p_1), p_2) & \text{if } \underline{i} = 1 \end{cases}. \tag{79}$$

Since $h(\text{int}(I_N)) \in \mathcal{I}_{\text{int},h}$, it follows that

$$\overline{d} = \begin{cases} d_{\underline{i}-1} & \text{if } \underline{i} > 1\\ p_2 & \text{if } \underline{i} = 1 \end{cases}, \tag{80}$$

and, from $(d_{N-1}, \overline{d}) \in \tilde{\mathcal{I}}_{\text{int},h} = \mathcal{I}_{\text{int}}$ (see (72)) and (37)–(38),

$$(d_{N-1}, \overline{d}) = \begin{cases} (d_{N-1}, d_{\underline{i}-1}) & \text{if } \underline{i} > 1\\ (d_{N-1}, p_2) & \text{if } \underline{i} = 1 \end{cases} = \text{int}(I_{N-1}). \quad \blacksquare$$

Appendix B
Proof of Proposition 5

We define N intervals $\tilde{I}_1, \ldots, \tilde{I}_N$ and prove that the properties (i)–(iv) hold for these. Let $m_1 := \bar{i}+1$ (> 1). We define recursively

$$\tilde{I}_{N-1} := h^{m_1}([d_{\bar{i}}, p_2]), \tag{81}$$

$$\tilde{I}_{i-_N 1} := h(\tilde{I}_i) \qquad \forall i \in \{1, \ldots, N-1\}, \tag{82}$$

where '$-_N$' is defined in (1).

We first show that (i)–(iii) hold for $\tilde{I}_{N-1}$. Notice that $\bar{i} \in \{1, \ldots, N-1\}$. We have

$$h([d_{\bar{i}}, p_2]) \underset{\text{Prop. 1 (i)}}{=} [h(d_{\bar{i}}), h(p_2)] = \begin{cases} [d_{\bar{i}-1}, h(p_2)] & \text{if } \bar{i} > 1\\ [p_1, h(p_2)] & \text{if } \bar{i} = 1 \end{cases}$$

$$\underset{(42)}{\subseteq} \begin{cases} [d_{\bar{i}-1}, d_{N-1}) & \text{if } \bar{i} > 1 \\ [p_1, d_{N-1}) & \text{if } \bar{i} = 1 \end{cases} \underset{\text{Lem. 1 (iii)}}{=} I_{\bar{i}-_N 1}. \tag{83}$$

From Proposition 4, it follows that, for all $i \in \{1, \dots, N\}$ and for all $m \in \mathbb{N}$,

$$h^m(I_i) \subseteq I_{i-_N m}, \tag{84}$$

$$h^m(\text{int}(I_i)) \subseteq \text{int}(I_{i-_N m}). \tag{85}$$

With this and (83), $h^{\bar{i}}([d_{\bar{i}}, p_2]) = h^{\bar{i}-1}(h([d_{\bar{i}}, p_2])) \subseteq h^{\bar{i}-1}(I_{\bar{i}-_N 1}) \subseteq I_{(\bar{i}-_N 1)-_N(\bar{i}-1)} = I_N$, and

$$\tilde{I}_{N-1} = h^{m_1}([d_{\bar{i}}, p_2]) = h^{\bar{i}+1}([d_{\bar{i}}, p_2]) \subseteq h(I_N) \underset{(39)}{=} h([p_1, d_{\underline{i}}))$$

$$\underset{\text{Prop. 1 (i)}}{=} \begin{cases} [h(p_1), d_{\underline{i}-1}) & \text{if } \underline{i} > 1 \\ [h(p_1), p_2) & \text{if } \underline{i} = 1 \end{cases}$$

$$\underset{\text{Prop. 2 (ii)}}{\subseteq} \begin{cases} (d_{N-1}, d_{\underline{i}-1}) & \text{if } \underline{i} > 1 \\ (d_{N-1}, p_2) & \text{if } \underline{i} = 1 \end{cases} \underset{\text{Lem. 1 (iv)}}{=} \text{int}(I_{N-1}) \subseteq I_{N-1}. \tag{86}$$

Thus, (ii) holds for $\tilde{I}_{N-1}$.

Property (i) can be seen as follows: $h([d_{\bar{i}}, p_2]) = [h(d_{\bar{i}}), h(p_2)]$ is closed. From (83) and Proposition 1, (i), it follows that h is continuous and strictly increasing on $h([d_{\bar{i}}, p_2])$. Similarly, by (84), $h^m([d_{\bar{i}}, p_2]) = h^{m-1}(h([d_{\bar{i}}, p_2])) \subseteq h^{m-1}(I_{\underline{i}-_N 1}) \subseteq I_{\underline{i}-_N m}$, $m \geq 1$; thus, h is continuous and strictly increasing on $h^m([d_{\bar{i}}, p_2])$. Since, for a continuous and strictly increasing function f and $a, b \in \mathbb{R}$, $f([a, b]) = [f(a), f(b)]$ (the image of a closed interval under f is a closed interval), $h^m([d_{\bar{i}}, p_2])$ is closed for any $m \geq 1$ and, in particular, for $m = m_1$.

To show (iii) for $\tilde{I}_{N-1}$, let $m_2 := N - m_1$ (≥ 0). We then get with (86), (84), and (38), $h^{m_2}(\tilde{I}_{N-1}) \subseteq h^{m_2}(I_{N-1}) \subseteq I_{(N-1)-_N m_2} = I_{\bar{i}} = [d_{\bar{i}}, p_2) \subseteq [d_{\bar{i}}, p_2]$. Property (iii) then follows by

$$h^N(\tilde{I}_{N-1}) = h^{m_1}(h^{m_2}(\tilde{I}_{N-1})) \subseteq h^{m_1}([d_{\bar{i}}, p_2]) \underset{(81)}{=} \tilde{I}_{N-1}.$$

Hence, we have shown that (i)–(iii) hold for $i = N - 1$. We next prove (i)–(iii) for $i \in \{1, \dots, N-2, N\}$ by induction.

Induction assumption (IA): (i)–(iii) are valid for some $i \in \{1, \dots, N-1\}$. Show that this implies that (i)–(iii) hold for $i -_N 1$.

Property (ii) holds since

$$\tilde{I}_{i-_N 1} \underset{(82)}{=} h(\tilde{I}_i) \underset{\text{IA (ii)}}{\subseteq} h(\text{int}(I_i)) \underset{(85)}{\subseteq} \text{int}(I_{i-_N 1}) \subseteq I_{i-_N 1}.$$

Since $\tilde{I}_i \subseteq I_i$ (IA (ii)), h is continuous and strictly increasing on $\tilde{I}_i$. Moreover, $\tilde{I}_i$ is closed (IA (i)). Together, this implies that the image under h, $\tilde{I}_{i-_N 1} = h(\tilde{I}_i)$, is also closed; hence, (i) is true.

Property (iii) can be seen to hold by

$$h^N(\tilde{I}_{i-_N 1}) \underset{(82)}{=} h^{N+1}(\tilde{I}_i) = h(h^N(\tilde{I}_i)) \underset{\text{IA (iii)}}{\subseteq} h(\tilde{I}_i) \underset{(82)}{=} \tilde{I}_{i-_N 1}.$$

This completes the proof of statements (i)–(iii).

To see statement (iv), take $I_i \in \mathcal{I}$ for any $i \in \{1, \dots, N\}$. Let $m_3 := i -_N \bar{i}$ (≥ 1). Then, from (84) and (38), $h^{m_3}(I_i) \subseteq I_{i-_N m_3} = I_{i-_N(i-_N\bar{i})} = I_{\bar{i}} = [d_{\bar{i}}, p_2) \subseteq [d_{\bar{i}}, p_2]$, and, with this and (81), $h^{m_1+m_3}(I_i) = h^{m_1}(h^{m_3}(I_i)) \subseteq h^{m_1}([d_{\bar{i}}, p_2]) = \tilde{I}_{N-1}$. Let $m_4 := (N -_N i) - 1$ ($0 \leq m_4 \leq N-1$). Then, with the preceding expression, (81), and (82), we get

$$h^{m_1+m_3+m_4}(I_i) = h^{m_4}(h^{m_1+m_3}(I_i)) \subseteq h^{m_4}(\tilde{I}_{N-1})$$
$$= \tilde{I}_{(N-1)-_N m_4} = \tilde{I}_{(N-1)-_N((N-_N i)-1)} = \tilde{I}_i. \tag{87}$$

Now, consider three different cases for i.

First case: $i = N$. Since $m_1 + m_3 + m_4 = (\bar{i}+1) + (N - \bar{i}) + (N-1) = 2N$, (iv) follows directly from (87).

Second case: $\bar{i} < i < N$. Since $m_1 + m_3 + m_4 = (\bar{i}+1) + (i - \bar{i}) + (N - i - 1) = N$, (87) reads $h^N(I_i) \subseteq \tilde{I}_i$, which implies (iv) as follows: $h^{2N}(I_i) = h^N(h^N(I_i)) \subseteq h^N(\tilde{I}_i) \subseteq \tilde{I}_i$.

Third case: $1 \leq i \leq \bar{i}$. Since $m_1 + m_3 + m_4 = (\bar{i}+1) + (i - \bar{i} + N) + (N - i - 1) = 2N$, (iv) follows from (87).

Appendix C
Proof of Lemma 2

Take $p \in [p_1, \bar{p} + \delta)$. Let $k \in \mathbb{N}^+$ such that $p, h(p), \dots, h^{k-1}(p) < \bar{p} + \delta$ (such a k exists since $p < \bar{p} + \delta$). Then, from (26),

$$h^k(p) = a^2\, h^{k-1}(p) + 1 > a^2\, h^{k-1}(p), \tag{88}$$

and, therefore,

$$h^k(p) > a^{2k}\, p. \tag{89}$$

Since $|a| > 1$, $\lim_{k\to\infty} a^{2k} p = \infty$. Hence, there exists an $m(p) \in \mathbb{N}^+$ such that (43) holds.

Now, we seek the largest possible integer $m(p)$ over all $p \in [p_1, \bar{p} + \delta)$, for which (43) holds. Since $h^k(p_1) \leq h^k(p)$ for all $p \in [p_1, \bar{p} + \delta)$ and $k \leq m(p)$, the greatest $m(p)$ such that (43) holds is $\bar{N} \in \mathbb{N}^+$ defined by

$$p_1, h(p_1), \dots, h^{\bar{N}-1}(p_1) < \bar{p} + \delta \quad \text{and} \quad h^{\bar{N}}(p_1) \geq \bar{p} + \delta. \tag{90}$$

Hence, $\bar{N}$ is independent of p, and $m(p) \leq \bar{N}$. From (89) and (90), it follows that

$$a^{2(\bar{N}-1)}\, p_1 < h^{\bar{N}-1}(p_1) < \bar{p} + \delta \underset{(p_1>0,\, a^2>0)}{\Rightarrow} a^{2\bar{N}} < a^2 \frac{\bar{p} + \delta}{p_1}.$$

References

[1] K.-D. Kim and P. Kumar, "Cyber-physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, May 2012.

[2] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, Jan. 2007.

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[4] S. Trimpe and R. D'Andrea, "An experimental demonstration of a distributed and event-based state estimation algorithm," in *Proc. of the 18th IFAC World Congress*, Milano, Italy, Aug. 2011, pp. 8811–8818.

[5] S. Trimpe, "Event-based state estimation with switching static-gain observers," in *Proc. of the 3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems*, Santa Barbara, CA, USA, Sep. 2012, pp. 91–96.

[6] S. Trimpe and R. D'Andrea, "Reduced communication state estimation for control of an unstable networked control system," in *Proc. of the 50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, FL, USA, 2011, pp. 2361–2368.

[7] ——, "The Balancing Cube: A dynamic sculpture as test bed for distributed estimation and control," *IEEE Control Systems Magazine*, vol. 32, no. 6, pp. 48–75, Dec. 2012.

[8] ——, "Event-based state estimation with variance-based triggering," in *Proc. of the 51st IEEE Conference on Decision and Control*, Maui, HI, USA, Dec. 2012, pp. 6583–6590.
[9] M. Lemmon, "Event-triggered feedback in control, estimation, and optimization," in *Networked Control Systems*, ser. Lecture Notes in Control and Information Sciences, A. Bemporad, M. Heemels, and M. Johansson, Eds. Springer Berlin / Heidelberg, 2011, vol. 406, pp. 293–358.
[10] O. C. Imer and T. Basar, "Optimal estimation with limited measurements," in *Proc. of the 44th IEEE Conference on Decision and Control and the European Control Conference*, Seville, Spain, Dec. 2005, pp. 1029–1034.
[11] L. Li, M. Lemmon, and X. Wang, "Event-triggered state estimation in vector linear processes," in *Proc. of the American Control Conference*, Baltimore, MD, USA, Jul. 2010, pp. 2138–2143.
[12] M. Rabi, G. V. Moustakides, and J. S. Baras, "Multiple sampling for estimation on a finite horizon," in *Proc. of the 45th IEEE Conference on Decision and Control*, San Diego, CA, USA, Dec. 2006, pp. 1351–1357.
[13] Y. Xu and J. P. Hespanha, "Estimation under uncontrolled and controlled communications in networked control systems," in *Proc. of the 44th IEEE Conference on Decision and Control and the European Control Conference*, Seville, Spain, Dec. 2005, pp. 842–847.
[14] R. Cogill, S. Lall, and J. P. Hespanha, "A constant factor approximation algorithm for event-based sampling," in *Proc. of the American Control Conference*, New York, NY, USA, Jul. 2007, pp. 305–311.
[15] J. Sijs and M. Lazar, "On event based state estimation," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, R. Majumdar and P. Tabuada, Eds. Springer Berlin / Heidelberg, 2009, vol. 5469, pp. 336–350.
[16] J. Sijs, M. Lazar, and W. P. M. H. Heemels, "On integration of event-based estimation and robust MPC in a feedback loop," in *Proc. of the 13th ACM international conference on hybrid systems: computation and control*, New York, NY, USA, 2010, pp. 31–40.
[17] J. Weimer, J. Araujo, and K. H. Johansson, "Distributed event-triggered estimation in networked systems," in *Proc. of the 4th IFAC Conference on Analysis and Design of Hybrid Systems*, Eindhoven, Netherlands, Jun. 2012, pp. 178–185.
[18] J. K. Yook, D. M. Tilbury, and N. R. Soparkar, "Trading computation for bandwidth: reducing communication in distributed control systems using state estimators," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 4, pp. 503–518, Jul. 2002.
[19] Y. Xu and J. P. Hespanha, "Optimal communication logics in networked control systems," in *Proc. of the 43rd IEEE Conference on Decision and Control*, Atlantis, Bahamas, Dec. 2004, pp. 3527–3532.
[20] H. Sandberg, M. Rabi, M. Skoglund, and K. H. Johansson, "Estimation over heterogeneous sensor networks," in *Proc. of the 47th IEEE Conference on Decision and Control*, Cancun, Mexico, Dec. 2008, pp. 4898–4903.
[21] L. Meier III, J. Peschon, and R. Dressler, "Optimal control of measurement subsystems," *IEEE Transactions on Automatic Control*, vol. 12, no. 5, pp. 528–536, Oct. 1967.
[22] H. Kushner, "On the optimum timing of observations for linear control systems with unknown initial state," *IEEE Transactions on Automatic Control*, vol. 9, no. 2, pp. 144–150, Apr. 1964.
[23] W. Zhang, M. P. Vitus, J. Hu, A. Abate, and C. J. Tomlin, "On the optimal solutions of the infinite-horizon linear sensor scheduling problem," in *Proc. of the 49th IEEE Conference on Decision and Control*, Atlanta, GA, USA, Dec. 2010, pp. 396–401.
[24] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Mineola, New York: Dover Publications, 2005.
[25] S. Bittanti, P. Colaneri, and G. De Nicolao, "The difference periodic Riccati equation for the periodic prediction problem," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 706–712, Aug. 1988.
[26] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.
[27] A. Molin and S. Hirche, "Structural characterization of optimal event-based controllers for linear stochastic systems," in *Proc. of the 49th IEEE Conference on Decision and Control*, Dec. 2010, pp. 3227–3233.
[28] ——, "On the optimality of certainty equivalence for event-triggered control systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 470–474, Feb. 2013.
[29] S. Elaydi, *An Introduction to Difference Equations*, 3rd ed., ser. Undergraduate Texts in Mathematics, S. Axler, F. W. Gehring, and K. A. Ribet, Eds. Springer New York, 2005.
[30] H. K. Khalil, *Nonlinear systems*, 3rd ed. Prentice Hall, 2002.
[31] W. G. Kelley and A. C. Peterson, *Difference equations: an introduction with applications*. San Diego, USA: Academic Press, Inc., 1991.
[32] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.
[33] E. T. Copson, *Metric spaces*. Cambridge University Press, 1968.

Sebastian Trimpe received the B.Sc. degree in General Engineering Science and the M.Sc. degree (Dipl.-Ing.) in Electrical Engineering from Hamburg University of Technology in 2005 and 2007, and the Ph.D. degree (Dr. sc.) in Mechanical Engineering from ETH Zurich in 2013.

Sebastian is currently a research scientist in the Autonomous Motion Department at the Max Planck Institute for Intelligent Systems in Tübingen, Germany. Previously, he was lecturer and postdoctoral researcher at ETH Zurich. In 2007, he stayed as a research scholar at the University of California at Berkeley. His research interests are in the area of control systems theory and design with emphasis on autonomous, networked, and learning systems. He is recipient of the General Engineering Award for the best undergraduate degree (2005), a scholarship from the German National Academic Foundation (2002 to 2007), and the triennial IFAC World Congress Interactive Paper Prize (2011).

Raffaello D'Andrea received the B.Sc. degree in Engineering Science from the University of Toronto in 1991, and the M.S. and Ph.D. degrees in Electrical Engineering from the California Institute of Technology in 1992 and 1997. He was an assistant, and then an associate, professor at Cornell University from 1997 to 2007.

While on leave from Cornell, from 2003 to 2007, he co-founded Kiva Systems, where he led the systems architecture, robot design, robot navigation and coordination, and control algorithms efforts. A creator of dynamic sculpture, his work has appeared at various international venues, including the National Gallery of Canada, the Venice Biennale, and the FRAC Centre. He is currently professor of Dynamic Systems and Control at ETH Zurich.