This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

Continuous Estimation Using Context-Dependent Discrete Measurements

Radoslav Ivanov, *Member, IEEE,* Nikolay Atanasov, *Member, IEEE,* Miroslav Pajic, *Member, IEEE,* James Weimer, *Member, IEEE,* George J. Pappas, *Fellow, IEEE,* and Insup Lee, *Fellow, IEEE*

***Abstract*—This paper considers the problem of continuous state estimation from discrete context-based measurements. Context measurements provide binary information as obtained from the system's environment, e.g., a medical alarm indicating that a vital sign is above a certain threshold. Since they provide state information, these measurements can be used for estimation purposes, similar to standard continuous measurements, especially when standard sensors are biased or attacked. Context measurements are assumed to have a known probability of occurring given the state; in particular, we focus on the probit function to model threshold-based measurements such as the medical-alarm scenario. We develop a recursive context-aware filter by approximating the posterior distribution with a Gaussian distribution with the same first two moments as the true posterior. We show that the filter's expected uncertainty is bounded when the probability of receiving context measurements is lower-bounded by some positive number for all system states. Furthermore, we provide an observability-like result – all eigenvalues of the filter's covariance matrix converge to 0 after repeated updates if and only if a persistence of excitation condition holds for the context measurements. Finally, in addition to simulation evaluations, we applied the filter to the problem of estimating a patient's blood oxygen content during surgery using real-patient data.**

I. Introduction

With the proliferation of sensing and computing technology, modern autonomous systems have access to a wealth of information when estimating their state. Given the recent improvements in machine learning, it is now possible to obtain high-level representations of this information. For example, if a robot detects a known building using image processing, the robot can conclude that it is near that building; similarly, if a medical device raises an alarm that a vital sign is above a certain threshold, it might be possible to conclude that the patient is in a critical state. Consequently, these discrete-valued context data can be viewed as measurements of (functions of) the system state, similar to conventional continuous sensors such as accelerometers or GPS (this notion is illustrated in Figure 1). Thus, context measurements can be used for state estimation both as a single source of information and in scenarios when some of the continuous sensors are noisy/biased (e.g., GPS in an urban environment [3] or medical sensors disrupted by moving artifacts [4]) or in security applications when some sensors might be attacked (e.g., the RQ-170 Sentinel drone that was captured in Iran [5] is believed to have had spoofed GPS [6]; if the drone had analyzed Iranian frequency modulation radio signals using natural language processing, it could have extracted context information that it is in Iran).

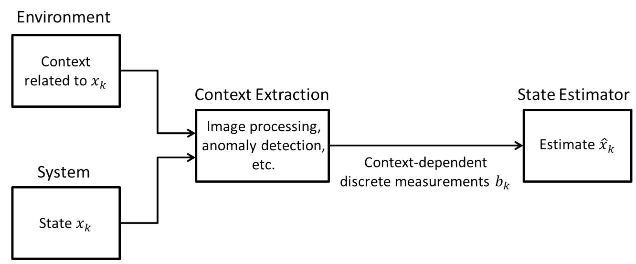


Fig. 1: General architecture of a system with access to context measurements.

In this paper, we develop a state estimation technique for linear systems with access to context measurements only. Context measurements are defined as discrete-valued data that have a known probability given the system state. Context measurements are especially useful when they represent low-level data that cannot be easily expressed as a function of the state (e.g., it is challenging to functionally map raw images to the robot's state). Thus, by using the probability distribution of context measurements given the state, one may use them for estimation in a rigorous manner. The probabilistic formulation makes sense intuitively – if a building is far from the robot and appears small in images, it might be recognized in some images only; if the building is nearby, we expect to recognize it in most images, i.e., the probability of receiving a context measurement would be high for states close to the building.

In this work, we are specifically interested in binary measurements as an important subclass of context measurements, i.e., each measurement is equal to 1 or -1 with a known probability given the state. Binary measurements capture a rich class of events that might occur during a system's operation. Examples include a medical device alarm that a vital sign exceeds a certain threshold (e.g., if the patient's oxygen

This work was supported in part by the NSF CNS-1505799, CNS-1652544 and CNS-1505701 grants, and the Intel-NSF Partnership for Cyber-Physical Systems Security and Privacy. This material is also based on research sponsored by the ONR under agreements number N00014-17-1-2012 and N00014-17-1-2504. Some of the preliminary results appeared in [1], [2].

R. Ivanov, J. Weimer, and I. Lee are with the Department of Computer and Information Science, University of Pennsylvania, 3330 Walnut Street, Philadelphia, PA 19104, e-mail: {rivanov, weimerj, lee}@seas.upenn.edu.

N. Atanasov is with the Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093, e-mail: natanasov@ucsd.edu.

G. J. Pappas is with the Department of Electrical and Systems Engineering, University of Pennsylvania, 3330 Walnut Street, Philadelphia, PA 19104, e-mail: pappasg@seas.upenn.edu.

M. Pajic is with the Department of Electrical and Computer Engineering, Duke University, 100 Science Dr, Durham, NC 27708, e-mail: miroslav.pajic@duke.edu.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

saturation is below a certain threshold, then the overall oxygen content (the state) must be below a certain threshold [2]) as well as occupancy grid mapping where a binary measurement is received as the robot gets close to an obstacle [7].

Estimation with context-based measurements was originally explored in radar target tracking where measurements also arrive irregularly and could be discrete [8] (refer to Section II for a discussion of related work). The models considered in this domain, however, are very general, which makes it challenging to derive exact theoretical results and instead leads to computationally expensive approximations. Approaches exist also for system identification with binary (but not random) observations [9] and for estimation with quantized measurements where measurements with known functional relation to the state (e.g., linear) are mapped to discrete sets, e.g., sign of innovations [10] or logarithmic quantizers [11].

In contrast with existing works, we develop a context-aware filter for linear systems with access to binary measurements. Unlike prior work, we assume no knowledge about the measurements other than their probability of occurring given the state. In particular, we focus on the probit function (i.e., the cumulative distribution function of the Gaussian distribution) in order to model the probability of getting context measurements given the state.[1] Since it resembles a step/sigmoid function, the probit function is well suited for modeling threshold-based context measurements – intuitively, the probability of getting a measurement is low when the state is well below the threshold and should rise as the state approaches/passes the threshold.

In our prior work [1], we presented the context-aware filter for the probit model by deriving the exact posterior distribution of the state given a context measurement. At the same time, it is not known how to compute the posterior for multiple context measurements since the integrals become intractable. As a result, we proposed to approximate the posterior distribution with a Gaussian distribution with the same first two moments as the true posterior. The approximating Gaussian distribution is then used as a prior for the next measurement, thus obtaining a recursive context-aware filter.

In this paper, we present theoretical analysis of the context-aware filter. We first show that the posterior distribution is unimodal, so that the Gaussian approximation is indeed justified. In addition, we show that, for a scalar system, the expected variance of the filter's estimates is bounded provided that the probability of receiving both a measurement of 1 and -1 is at least some positive number η. This result is similar to a corresponding fact about Kalman filtering with intermittent observations [12] in the sense that the system needs to perform "useful" updates often enough in order to keep the uncertainty bounded. Generalizing this result to multidimensional systems, however, is challenging due to the fact that we aim to estimate continuous variables using discrete measurements only; at the same time, the same intuition could be used to prove a similar claim in the multidimensional case as well.

To provide further intuition about the filter's performance in the multidimensional case, we show convergence results about systems with no dynamics. We show that the eigenvalues of the filter's covariance matrix converge to 0 if and only if a persistence-of-excitation condition holds for the context measurements. This result is the context equivalent to an observability claim in a standard linear system – intuitively, if there exist context measurements that observe all states, then the uncertainty decreases over time. Furthermore, we show that as the eigenvalues of the covariance matrix converge to 0, the expressions for the moments of the Gaussian approximations converge to a form similar to the Newton method [13], which suggests that the estimates likely converge to the true state, since the posterior distribution is unimodal. This result provides a parallel with the widely used Expectation Propagation [14] algorithm where similar Gaussian approximations are employed – thus, the results presented in this paper might be of interest to the machine learning community as well.

Finally, we evaluate the context-aware filter both in simulation and on real-patient data collected from the Children's Hospital of Philadelphia (CHOP). We first show the evolution of the estimates for a system with no dynamics in order to illustrate the saw-shaped nature of the estimation curve induced by binary measurements. In addition, we simulate a moving system in order to illustrate a case in which the estimator does converge for moving systems as well. Finally, we apply the filter to the problem of estimating a patient's blood oxygen (O_2) content during surgery. Since the O_2 content cannot be measured non-invasively, we use context measurements extracted from different medical device data to perform estimation. The results indicate that adding context reduces the estimation error by about 20%, on average.

The remainder of this paper is organized as follows. Section II provides a discussion on related work in several research communities. Section III formulates the problem addressed in this work, and Section IV presents the context-aware filter. The convergence analysis of the filter is shown in Section V. We evaluate the filter's performance in Section VI (in simulation) and in Section VII (on real data). Finally, Section VIII provides concluding remarks.

II. Related Work

The concept of context-aware filtering has appeared in different forms in several research communities. As mentioned in Section I, there exist target tracking approaches for filtering with both discrete and continuous measurements, e.g., the probability hypothesis density (PHD) filter [8]. Other non-linear filters have been developed as well, such as the hybrid density filter (HDF) [15], the set-membership filter [16], and the assumed density filter (ADF) [17] (the context-aware filter is a type of ADF for which we can compute the moments of the posterior distribution). Due to their generality, however, these filters do not provide strong theoretical guarantees about specific classes of non-linear systems; in contrast, by focusing on a specific class of non-linear measurements, we can derive a closed-form filter with strong theoretical properties.

Context measurements are also similar to quantized measurements in that they are discrete-valued [11], [10]. Quantized measurements are different, however, because they are

[1] In prior work [1], we also considered a second class of probability of detection functions, namely inverse-exponential functions. In the interest of space, however, that discussion is not included here.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

derived from standard continuous measurements whereas context measurements are only related to the state through the probability of detection. System identification with binary measurements [9] has also been investigated although no approaches exist for the probabilistic setting in our paper.

Context-aware filtering is also similar to Kalman filtering with intermittent observations [12], [18] and unreliable links [19], [20], [21], [22] in that measurements arrive irregularly, and the frequency of measurement arrivals affects the filter's performance. Related to this is the area of sensor scheduling where different sensors are used at different times so as to minimize interference or power consumption [23], [24], [25]. Yet another similar problem has been considered in wireless sensor networks where sensors are deployed over a large area such that the receipt of each sensor's measurement could be considered a context measurement [26], [27].

Due to their discrete nature, context measurements can also be modeled with hybrid systems [28], where different modes contain different models of context measurements. Such models include Markov chain switching [29], [30], deterministic switching [31], [32] and other more general models [33]. However, due to their complexity, all of these approaches rely on approximations in order to perform the estimation task.

Different notions of context are also used in robotics for the purpose of localization and mapping [34] by using scene categorization [35] and object class information [36], [37]. However, these papers do not provide theoretical guarantees for their approaches. The work that is closest in its setup and assumptions to our paper addresses the problem of indoor localization by using both continuous and discrete measurements [37]; however, the particle filter used to combine the two types of measurements does not provide theoretical guarantees for a finite set of particles and may suffer from particle deprivation problems in high-dimensional spaces. Finally, context-aware filtering is also related to Gaussian process classification [38] since the objective is to learn a continuous probability distribution from discrete data. In particular, the EP algorithm [14] is similar to the context-aware filter in that posteriors are approximated with Gaussian distributions as well; however, no convergence results exist for EP.

III. Problem Formulation

This section presents the system model used in this paper, including the probit context measurement model. The precise problem statement is provided at the end of the section.

A. System Model

Consider a linear discrete-time system of the form

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

where $x \in \mathbb{R}^n$ is the system state, $x_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$, $w_k \sim \mathcal{N}(0, Q)$ is Gaussian process noise, and A_k is a matrix of appropriate dimensions describing the system dynamics.

Instead of the classical continuous sensors, the system considered in this paper only has access to context sensors.[2] Context sensors provide binary information about the system's context; examples include detecting nearby objects with known positions on a map or a vital sign exceeding a certain predefined threshold. At each time k, a measurement b_k is received that is equal to 1 if a detection occurs and -1 otherwise.[3] We assume that b_k is equal to 1 with a known probability given the state, denoted by $p_k^d(b_k \mid x_k)$, i.e.,

$$b_k = \begin{cases} 1 & w.p. \quad p_k^d(b_k \mid x_k) \\ -1 & w.p. \quad 1 - p_k^d(b_k \mid x_k). \end{cases} \tag{2}$$

As noted in Section I, p_k^d is close to 1 when the system is in a state that is highly correlated with receiving a context measurement (e.g., a robot is close to a building). Note that p_k^d is time-varying, i.e., different binary measurements may be received at different times. It is assumed that, conditioned on the state, context measurements are mutually independent.

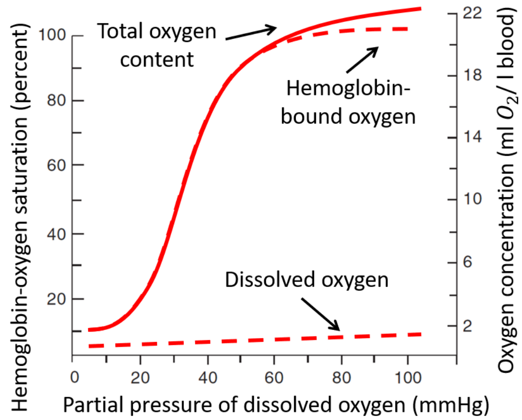


Fig. 2: Most of the O_2 in the blood is bound to hemoglobin.

B. Context Measurement Model

As argued in Section I, we use the probit function to model the probability of detection of context measurements [39]:

$$p_k^d(b_k \mid x_k) = \Phi((v_k^T x_k + a_k) b_k), \tag{3}$$

where Φ is the cumulative distribution function (cdf) of the standard Normal distribution, $v_k \in \mathbb{R}^n$ is a vector of known parameters, and $a_k \in \mathbb{R}$ is a known parameter offset. Note that $p_k^d(b_k = 1 \mid x_k) = 1 - p_k^d(b_k = -1 \mid x_k)$ due to the rotational symmetry of Φ, i.e., $\Phi(-x) = 1 - \Phi(x)$. We assume there is a finite set of size C of context weights and offsets $\mathcal{V} = \{(v^1, a^1), \ldots, (v^C, a^C)\}$.

Due to its step-like shape, the probit function is well suited for modeling threshold-based events such as medical alarms. Consider the problem of estimating the patient's O_2 content ($C_a O_2$); as shown in Figure 2, most of the O_2 is bound to hemoglobin. Although the precise mapping from hemoglobin-oxygen saturation ($S_p O_2$) to $C_a O_2$ is unknown (and varies

[2] All results in this paper also hold in the addition of classical measurements of the form $y_k = C x_k$ (plus Gaussian noise). To keep the presentation simple, however, we focus on the case with context measurements only.

[3] Our framework can handle more than one binary measurement by repeated updates. We make the one-measurement assumption to simplify notation.

across patients), if C_aO_2 is below a threshold t_s, then one also expects to see a measurement of S_pO_2 below a threshold t_m. Thus, we can introduce a context measurement b_k that is equal to 1 if $S_pO_2 > t_m$ and -1, otherwise.

To relate b_k to the state (C_aO_2), note that as C_aO_2 becomes much smaller than t_s, it becomes more likely for b_k to be -1; conversely, if C_aO_2 is greater than t_s, it is very unlikely for b_k to be -1. The probit function is ideal for capturing such a scenario: the probability of $b_k = 1$ is close to 0 for low values of C_aO_2 and approaches 1 as C_aO_2 rises above t_s. The parameters in the probit function should be chosen based on the following considerations: since v_k determines the slope of the step-like response in the probit function, v_k should be large if the relationship between t_s and t_m is precise (e.g., if $S_pO_2 < t_m$, then necessarily $C_aO_2 < t_s$) and should be smaller if some false positives are expected; since a_k determines the threshold where the step response begins, a_k should be set to $-v_k t_s$ (in the one-dimensional case) to ensure the probability rises quickly as the threshold is crossed.

C. Problem Statement

Problem: Given the system defined in (1)-(3) and a prior probability density function (pdf) $p_{k|k}(x) = p(x \mid b_{0:k})$ the goal is to compute (and analyze) the posterior density

$$p_{k+1|k+1}(x) := p(x \mid b_{0:k+1}).$$

IV. Context-Aware Filter

The problem formulation in Section III naturally leads to a Bayesian filter of the form:

$$\begin{aligned}\textbf{Predict: } & p_{k+1|k}(x) = \int p^f_{k+1|k}(x \mid z) p_{k|k}(z) dz, \\ \textbf{Update: } & p_{k+1|k+1}(x) = \xi_{k+1} p^d_k(b_{k+1} \mid x) p_{k+1|k}(x),\end{aligned} \tag{4}$$

where $p^f_{k+1|k}(x_{k+1} \mid x_k)$ is the conditional pdf of the state at time $k+1$ given the state at time k and ξ_{k+1} is a constant [40].

Equation (4) is impossible to derive in closed form for arbitrary dynamics and observation models (with the exception of the linear Gaussian case, which leads to the Kalman filter). As discussed in Section II, multiple approximation approaches with different assumptions exist, such as the ADF, PHD filter, and the HDF. Due to their generality, all of these approaches rely on approximations when computing their estimates.

That is why, in this paper we focus on a specific observation model (i.e., the probit model defined in (3)) and derive the exact posterior distribution after the update with a binary measurement. At the same time, developing a closed-form recursive filter is not straightforward, since the posterior distribution is no longer Gaussian. As we argue below, however, a Gaussian distribution with the same mean and covariance matrix is a good approximation for the resulting posterior distribution since the true posterior is unimodal as well.

The next subsections present the recursive context-aware filter, assuming the prior $p_{k-1|k-1}$ is a **Gaussian distribution with mean** $\mu_{k-1|k-1}$ **and covariance matrix** $\Sigma_{k-1|k-1}$.

A. Predict

The predict phase is the classical Kalman filter prediction:

$$\begin{aligned}p_{k|k-1}(x) &= \int \phi(x; A_{k-1}z, Q)\phi(z; \mu_{k-1|k-1}, \Sigma_{k-1|k-1})dz \\ &= \phi(x; A_{k-1}\mu_{k-1|k-1}, A_{k-1}\Sigma_{k-1|k-1}A^T_{k-1} + Q) \\ &= \phi(x; \mu_{k|k-1}, \Sigma_{k|k-1}),\end{aligned}$$

where $\phi(x; \mu, \Sigma)$ denotes the pdf of a Gaussian distribution with mean μ and covariance matrix Σ.

B. Update

The posterior distribution after the receipt of a binary measurement b_k is shown in Proposition 1 below (all proofs are given in the Appendix).

Proposition 1: Upon receipt of a discrete measurement $b_k \in \{-1, 1\}$, the discrete update is as follows:

$$p_{k|k}(x) = \frac{\Phi((v_k^T x + a_k)b_k)\phi(x; \mu_{k|k-1}, \Sigma_{k|k-1})}{Z_k}, \tag{5}$$

where

$$Z_k = \Phi\left(\frac{(v_k^T \mu_{k|k-1} + a_k)b_k}{\sqrt{v_k^T \Sigma_{k|k-1} v_k + 1}}\right).$$

Approximation: We approximate the posterior distribution in (5) with a Gaussian distribution with the same mean and covariance matrix.

The posterior distribution in (5) is no longer Gaussian. In fact, related work [41], [42] has shown that the posterior is not Gaussian in multiple truncation scenarios, e.g., with infrequent measurement transmissions. In such cases, it might be possible to develop filters for skewed normal distributions. However, a Gaussian still seems to be a good approximation for (5). In particular, as shown in Proposition 2 below, the distribution in (5) is log-concave; log-concavity, in turn, implies unimodality, as discussed in Corollaries 1 and 2.

In addition, despite the filter's discrete nature, the posterior distribution in (5) is not the result of a truncation process but is actually smooth (infinitely differentiable, in fact). This suggests that no individual measurement can introduce large skewness to either side. Finally, the pdf in (5) is computed numerically in Section VI-C (for multiple updates); the results provide strong evidence that the posterior is sufficiently symmetric so that a Gaussian approximation is justified. Thus, we approximate the posterior in (5) with a Gaussian with the same mean and covariance matrix as the distribution in (5) – these quantities are computed in Proposition 3 below.

Proposition 2: The distribution in (5) is log-concave, i.e., the function $g(x) = \ln(p_{k|k}(x))$ is concave.

Corollary 1 ([43]): In one dimension, the distribution in (5) is **unimodal**, i.e., there exists a point x^* such that $p_{k|k}(x)$ is increasing for $x \leq x^*$ and $p_{k|k}(x)$ is decreasing for $x \geq x^*$.

Corollary 2 ([43]): In many dimensions, the distribution in (5) is **star-unimodal** (a random variable $X \in \mathbb{R}^n$ is said to have a star-unimodal distribution if for every bounded non-negative Borel measurable function f on $\mathbb{R}^n$, $t^n\mathbb{E}[f(tX)]$ is non-decreasing for $t \in [0, \infty)$).[4]

Proposition 3: The mean of the distribution in (5) is:

$$\mu_{k|k} = \mu_{k|k-1} + \Sigma_{k|k-1}v_k(v_k^T\Sigma_{k|k-1}v_k + \chi_k)^{-1}b_k, \quad (6)$$

where

$$\chi_k = \frac{\sqrt{v_k^T\Sigma_{k|k-1}v_k + 1} - v_k^T\Sigma_{k|k-1}v_k\alpha(M_k)}{\alpha(M_k)} \quad (7)$$

$$\alpha(x) = \phi(x; 0, 1)/\Phi(x) \quad (8)$$

$$M_k = \frac{(v_k^T\mu_{k|k-1} + a_k)b_k}{\sqrt{v_k^T\Sigma_{k|k-1}v_k + 1}}. \quad (9)$$

The covariance matrix of the distribution in (5) is:

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \Sigma_{k|k-1}v_k(v_k^T\Sigma_{k|k-1}v_k + \gamma_k)^{-1}v_k^T\Sigma_{k|k-1} \quad (10)$$

where

$$\gamma_k = \frac{(1 - h(M_k))\, v_k^T\Sigma_{k|k-1}v_k + 1}{h(M_k)} \quad (11)$$

$$h(x) = \alpha(x)(x + \alpha(x)). \quad (12)$$

***Remark**:* The context-aware filter is similar to Kalman filtering with intermittent observations [12] in that measurements arrive in a stochastic manner. Thus (10) resembles a standard Riccati equation (update), where the non-linear term γ_k could be considered as the equivalent of measurement noise.

Note also that the functions α and h defined in (8) and (12), respectively, have been studied extensively in the statistics community. The ratio α is known as the inverse Mills ratio; some properties of the inverse Mills ratio that are used throughout this paper are summarized below.

***Definition**:* The inverse Mills ratio is defined as the ratio of the pdf and cdf of a standard Normal distribution, respectively:

$$\alpha(x) = \phi(x; 0, 1)/\Phi(x).$$

Proposition 4 ([44]): The following statements are true about the inverse Mills ratio:

1) $h(x) := -\alpha'(x) = \alpha(x)(x + \alpha(x))$
2) $0 < h(x) < 1, \forall x \in \mathbb{R}$
3) $h'(x) < 0, \forall x \in \mathbb{R}$.

***Remark**:* Since $0 < h(x) < 1$, we can conclude that $\gamma_k > 1$.

V. Convergence Properties

In this section we analyze the convergence properties of the context-aware filter. Since the task is to estimate a continuous variable using only discrete measurements, proving convergence is hard in general, especially given the random and time-varying nature of the filter. Ideally, one could hope to prove that the expected covariance matrix is bounded under some conditions on the initial condition and the probability of measurement arrivals (similar to Kalman filtering with intermittent observations [12]). However, the random non-linear term γ_k in the covariance matrix update in (10) makes it challenging to analyze the system when dynamics are also considered since γ_k cannot be upper-bounded in general (as shown in Proposition 4, the function h can be arbitrarily close to 0). Such an upper bound can be derived in the special case of a scalar system as shown in the next subsection.

To provide further intuition about the filter's convergence, we also show results for a non-moving system. In particular, in Subsections V-B and V-C we provide an observability-like claim for the filter, i.e., the eigenvalues of the covariance matrix converge to 0 if and only if a persistence-of-excitation condition is true for the weight vectors v_k over time. Furthermore, we show that, as the eigenvalues of the covariance matrix converge to 0, the discrete update of the filter converges to a Newton-Method-like step, which is an intuitive result given that the filter approximation matches the first two moments of the true posterior distribution.

A. Bounded Variance for a Scalar System

In this section we analyze conditions that result in a bounded variance of the context-aware filter given a scalar system:

$$x_{k+1} = ax_k + w_k, \quad (13)$$

where $x_k, a \in \mathbb{R}$, and $w_k \sim \mathcal{N}(0, q)$.

First note that the update in (10) looks like a standard Riccati equation, except for the non-linear term γ_k. Thus, one way to show that the context-aware filter's variance is bounded is by providing an upper bound on γ_k such that (10) is bounded (with some positive probability) by a standard Riccati equation. In such a case, our problem can be reduced to Kalman filtering with intermittent observations [12], and we can use some of the known facts for that scenario.

One case in which γ_k can be bounded (with positive probability) is when the probability of receiving both a measurement of 1 or -1 is at least some positive number η. In such a case, γ_k can be upper-bounded (with probability at least η) by $((1-h(0))v_k\sigma_k v_k+1)/h(0)$ by using the fact that $h'(x) < 0$ for all x. This condition leads to the following result, similar to a result from Kalman filtering with intermittent observations.

Theorem 1: Consider the system in (13) and suppose that, for all x_k, $p_k^d(b_k \mid x_k) \geq \eta$ for $b_k = \pm 1$. Then there exists some $\eta_c \in [0, 1)$ such that

$$\forall \sigma_0, \mathbb{E}[\sigma_k] \leq M_{\sigma_0}, \text{for } \eta_c < \eta \leq 1,$$

where M_{σ_0} is a constant that depends on the initial condition. Theorem 1 says that the filter's expected uncertainty is bounded if the probability of receiving "useful" measurements is sufficiently high (by "useful" we mean that a measurement can be both 1 or -1 with probability at least η such that receiving the measurement does provide significant information). This result makes sense intuitively – if the system is moving away from all available context measurements (i.e., if v^Tx+a is very large in absolute value for all $(v, a) \in \mathcal{V}$), we cannot expect to be able to estimate the state; conversely, if context measurements are available throughout the system's execution, then the filter's uncertainty should be low.

[4]While there is a standard definition of unimodality in one dimension, many definitions exist in multiple dimensions [43].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

The proof of Theorem 1 does not generalize immediately to the multidimensional case, as the bound on γ_k does not lead to a standard-Riccati-equation bound on the expected covariance matrix. The multidimensional modified Riccati equation effectively has a time-varying covariance matrix that is difficult to bound; establishing the convergence of such a filter is an open problem in control theory and is part of future work. At the same time, we believe the same intuition holds for the multidimensional case as well.

B. Covariance Matrix Convergence for Non-Moving System

While we cannot bound the filter's expected uncertainty in the multidimensional case, we provide such a result in the special case of a non-moving system. Estimation for non-moving systems has interesting applications as well, e.g., the robotics mapping problem where a robot with a known position attempts to locate all (non-moving) obstacles on the map by receiving binary measurements when objects are detected. We show that for a system with no dynamics, the eigenvalues of the covariance matrix converge to 0 if and only if a persistence-of-excitation condition (formalized below) is true for the weight vectors v_k over time.

To simplify notation and since no dynamics predictions are performed in this section, we drop the prediction notation in the rest of this section (i.e., we write Σ_k instead of $\Sigma_{k|k} = \Sigma_{k+1|k}$). Before presenting the main result of this subsection, we first describe the behavior of the covariance matrix after multiple binary updates, as presented in the following lemma.

Lemma 1: After applying N updates at time k, the covariance matrix update from (10) can be written as:

$$\Sigma_{k+N} = \Sigma_k - \Sigma_k V_k^T (V_k \Sigma_k V_k^T + \Gamma_k)^{-1} V_k \Sigma_k, \qquad (14)$$

where $V_k = [v_{k+1}, \ldots, v_{k+N}]^T$, $[\Gamma_k]_{(i,j)} = \gamma_{k+i}$ if $i = j$ and $[\Gamma_k]_{(i,j)} = 0$ otherwise.

The update in Lemma 1 is similar to a standard Riccati equation (without the dynamics elements). Thus, it is not surprising that convergence of the covariance matrix depends on similar conditions on the matrix V_k as for a C_k matrix in a standard linear system. One such property is the widely used persistence of excitation [45].

Definition *(Persistence of Excitation):* The sequence of context weights and offsets, (v_k, a_k), is **persistently exciting** if there exist n linearly independent weight vectors with corresponding offsets $\mathcal{P} = \{(v^1, a^1), \ldots, (v^n, a^n)\}$ that appear infinitely often, i.e., for every k, there exists $l_k \in \mathbb{N}$ such that

$$\forall (v^i, a^i) \in \mathcal{P}, \exists t \in \{k, \ldots, k + l_k\} \text{ s.t. } (v_t, a_t) = (v^i, a^i).$$

Persistence of excitation is a standard assumption in estimation and system identification [45].[5] Intuitively, it means that there exists a set of context measurements that are received infinitely often such that their corresponding weights span $\mathbb{R}^n$.[6] The offsets are also important because even if the same weights repeat over time, the change of offsets might still affect the probability of receiving new context measurements.

Theorem 2: Suppose the system has no dynamics (i.e., $A_k = I$, the identity matrix, and $Q = 0$). Let $\lambda_k^j > 0$ be the eigenvalues of Σ_k. Then $\lambda_k^j \xrightarrow{a.s.} 0$ as $k \to \infty$ if and only if (v_k, a_k) is persistently exciting.

Theorem 2 is essentially an observability result. It suggests that if some states are not observed through binary measurements, then the uncertainty about those states does not decrease over time. If all states are observed, however, then the uncertainty is reduced in a manner similar to the standard Kalman filter with a persistently exciting C_k matrix.

Even if the covariance matrix converges to zero, it is not clear whether the filter's estimates converge to the true state. However, as shown in Section VI, simulations suggest that the estimates do converge to the true state. Furthermore, similar convergence results exist for the EP algorithm (which also contains a Gaussian approximation), namely 1) EP converges to the true state for strongly log-concave observation models [46] (the probit model is log-concave but is not strongly log-concave) and 2) in the limit, EP has a fixed point at the true state if the observation model has bounded derivatives [47] (true for the probit model). Thus, it is likely that the context-aware filter's mean also converges to the true state but we leave proving this result for future work.

C. Convergence of "Site" Approximations

In an effort to better understand the asymptotic behavior of the context-aware filter for systems with no dynamics, in this subsection we analyze the effect of a single update in the limit. In particular, we show that as more data is available, discrete updates converge to a Newton-Method-like step (this result is similar to a recent result about the limit behavior of EP [47]).

Definition: The Newton Method for finding the minimum of a twice-differentiable function f is computed as follows: given the previous iteration point x_n, the next step is [13]

$$x_{n+1} = x_n - \left[f''(x_n)\right]^{-1} f'(x_n).$$

The significance of this property is that the Newton Method converges to the optimal value (i.e., the peak of the distribution) of concave or quasi-concave functions. Since the posterior distribution in (5) is log-concave (i.e., quasi-concave), there is strong evidence to believe that the context-aware filter does indeed converge to the true state.

Each update of the context-aware filter could be viewed as a Gaussian approximation of the observation model itself (i.e., of the probit model). Specifically, the posterior Gaussian approximation could be considered as a Gaussian distribution that resulted from an update in which the observation model was also a Gaussian distribution with the appropriate parameters (also known as a "site" approximation in machine learning).

Definition *(Site Approximation):* Given a Gaussian prior $\phi(x; \mu_{k-1}, \Sigma_{k-1})$ and a binary update with observation model $\Phi((v_k^T x + a_k) b_k)$, a site approximation is a Gaussian distribution $p^s(x) := \phi(x; \mu^s, \Sigma^s)$ such that the distribution (normalized by the constant β)

$$p^G(x) = \beta \phi(x; \mu_{k-1}, \Sigma_{k-1}) \phi(x; \mu^s, \Sigma^s)$$

[5] The definition used in our paper is a special case of standard definitions since we have a finite set of context weights.

[6] Persistence of excitation does not require the received context measurements to take on a specific value, i.e., they can be either -1 or 1. Intuitively, the definition only requires the same classifiers to run infinitely often.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

has the same mean and covariance matrix as the true posterior

$$p_{k|k}(x) = \frac{1}{Z_k}\Phi((v_k^T x + a_k)b_k)\phi(x;\mu_{k-1},\Sigma_{k-1}).$$

Site approximations are easily computed when we consider the natural parameters of the distribution. Suppose the prior distribution is $\phi(x;\Omega_{k-1}^{-1}\omega_{k-1},\Omega_{k-1}^{-1})$, where $\Omega_{k-1} = \Sigma_{k-1}^{-1}$ and $\omega_{k-1} = \Omega_{k-1}\mu_{k-1}$ are the prior's information matrix and mean, respectively. Similarly, suppose the posterior Gaussian approximation is $\phi(x;\Omega_k^{-1}\omega_k,\Omega_k^{-1})$. Then the parameters of the site approximation $\phi(x;(\Omega_k^s)^{-1}\omega_k^s,(\Omega_k^s)^{-1})$ are [48]:

$$\Omega_k^s = \Omega_k - \Omega_{k-1} \tag{15}$$

$$\omega_k^s = \omega_k - \omega_{k-1}. \tag{16}$$

The site approximation abstraction is useful as it allows us to reason about the "contribution" of each update.

Theorem 3: Suppose the prior is $\phi(x;\Omega_k^{-1}\omega_k,\Omega_k^{-1})$ (where $\Omega_k = \Sigma_k^{-1}$ and $\omega_k = \Omega_k\mu_k$). After performing an update in the context-aware filter, the natural parameters of the site approximation are:

$$\Omega_{k+1}^s = v_{k+1}\gamma_{k+1}^{-1}v_k^T \tag{17}$$

$$\omega_{k+1}^s = \Omega_{k+1}^s\mu_k + (I + L_{k+1})v_{k+1}N_{k+1}^{-1}b_{k+1}, \tag{18}$$

where

$$N_{k+1} = v_{k+1}^T\Sigma_k v_{k+1} + \chi_{k+1}$$

$$L_{k+1} = v_{k+1}\gamma_{k+1}^{-1}v_{k+1}^T\Sigma_k.$$

Corollary 3: Suppose the system has no dynamics (i.e., $A_k = I$, the identity matrix, and $Q = 0$). If (v_k, a_k) is persistently exciting, then the natural parameters of the site approximations converge to

$$\Omega_{k+1}^s \xrightarrow{a.s.} \psi_{k+1}^{''}(\mu_k) \tag{19}$$

$$\omega_{k+1}^s \xrightarrow{a.s.} \Omega_{k+1}^s\mu_k - \psi_{k+1}^{'}(\mu_k), \tag{20}$$

where ψ_{k+1} is the negative log-likelihood of the measurement b_{k+1}, i.e.,

$$\psi_{k+1}(x) = -\ln(\Phi((v_{k+1}^T x + a_{k+1})b_{k+1})).$$

Remark: Since $\Omega_{k+1}^s\mu_{k+1}^s = \omega_{k+1}^s$, we can conclude that $\mu_{k+1}^s \xrightarrow{a.s.} \mu_k - [\psi_{k+1}^{''}(\mu_k)]^{-1}\psi_{k+1}^{'}(\mu_k)$. This is not the same as the Newton Method since it contains the site approximation mean instead of the posterior distribution mean. Yet, it shows that the site approximations themselves behave as a Newton Method update that is added to the prior mean.

The significance of Corollary 3 is that the Newton Method converges to the minimal (maximal) point of a log-convex (-concave) function. Although the site approximations are not identical to the Newton Method (since the $\psi_{k+1}(x)$ functions change over time), they do perform a Newton Method update at each time step. In turn, a Newton Method behavior implies that the site approximations converge to the Canonical Gaussian Approximation (CGA) [46], i.e., the Gaussian distribution whose mean is the maximizer of the true observation model's probability distribution and whose covariance matrix is the Hessian at that maximum. Finally, it is known that CGA's converge almost surely to a large class of posterior distributions, e.g., as shown by the Bernstein-von Mises Theorem [49]. Thus, Corollary 3 presents strong evidence to believe that the context-aware filter does indeed converge to the mean of the true posterior distribution. The next Section presents several simulation scenarios in support of this claim as well.

VI. Simulation Evaluation

We evaluate the context-aware filter both in simulation and on real data collected from CHOP. In this section we provide evaluation in simulation. The next section presents the application of the context-aware filter to the problem of estimating the blood oxygen content during surgery.

A. System with No Dynamics

We first evaluate the filter on a system with no dynamics, in order to illustrate the significance of Theorem 2. Figure 3 shows the filter's evaluation on a scalar system with a constant state $x_k = 3$ and with access to one context measurement with parameters $v_k = 1$ and $a_k = -5$. The initial condition is $\mu_0 = 1$, $\Sigma_0 = 2$. Figure 3c shows the evolution of the covariance for 10 runs of the system; as expected, the covariance converges to 0 for each one, thus ensuring the convergence of the filter overall. Figure 3b shows the estimation errors for the same 10 runs; the estimates are close to the true state, although some estimates converge more slowly due to different random realizations of the measurements. Finally, Figure 3a shows the toothed shape of the estimates for an example run, with discrete jumps as new context measurements are incorporated.

B. System with Unstable Dynamics

In the second simulation, we evaluate the filter's performance on an unstable system. The system dynamics are:

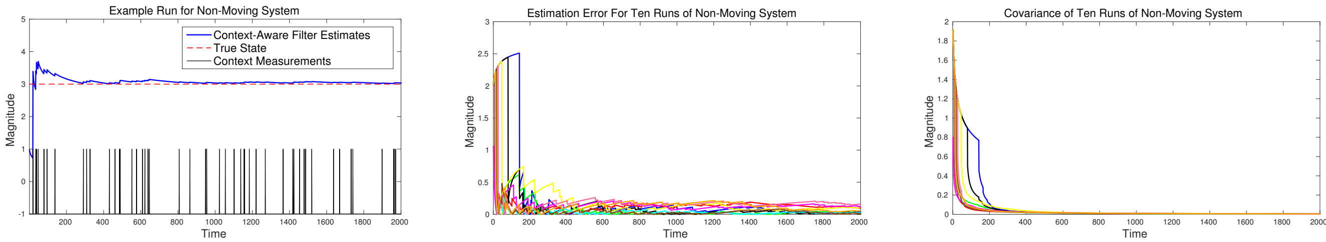
$$x_{k+1} = \begin{bmatrix} 1.01 & 0 \\ 0 & 1.01 \end{bmatrix} x_k + w_k,$$

where $w_k \sim \mathcal{N}(0, 0.001I)$ and $x_0 = [1 \quad 1]^T$.[7] 24 context measurements are received at each time, 12 with weights $v_{k,1} = [0 \quad 1]^T$ and 12 with weights $v_{k,2} = [1 \quad 0]^T$; the 12 offsets a_k are decreased linearly from 0 to -240 (i.e., they provide rough information as to whether each state is between 0 and 20, 20 and 40, etc.).

Figure 4 shows the results of the simulation. Figure 4a shows that the filter tracks the state very well after the initial period of uncertainty. The total number of context measurements equal to 1 at each step are shown in Figure 4b; as can be seen in the figure, eventually the system crosses almost all 24 context thresholds. In addition, we observe similar trends as in Figure 3, i.e., the estimates track the real system well after the initial period of uncertainty (Figure 4c), and the trace of the covariance matrix (Figure 4d) converges over time. The spikes in the trace of the covariance matrix around step 200 are due to the fact that the system receives the same context measurements around steps 150-230; once more context thresholds are crossed, the filter's uncertainty decreases again. These results suggest that the filter does converge over time (given certain observability-like conditions) and is likely asymptotically unbiased.

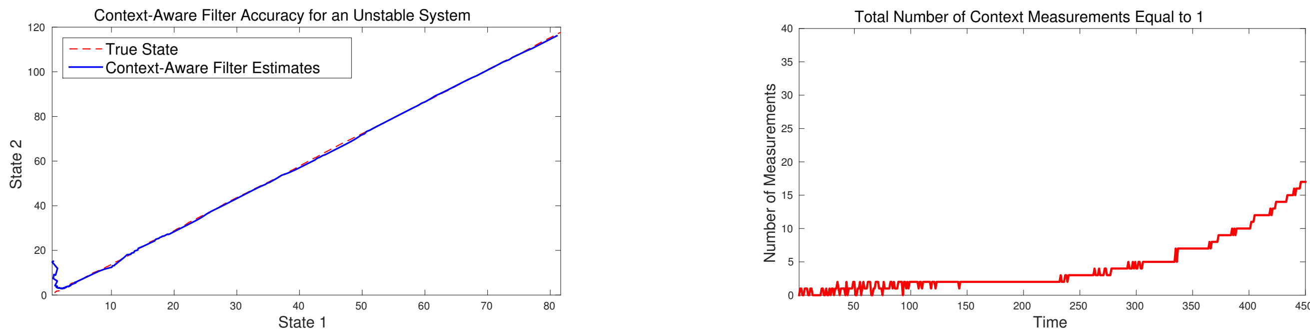
[7] Systems with larger-eigenvalue dynamics were tested as well with similar results; the system used in this section was chosen for visualization purposes.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control



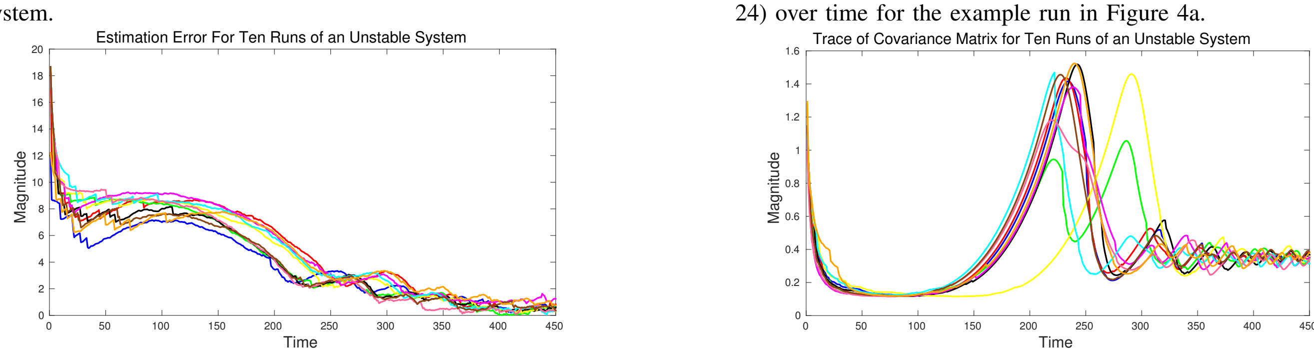
(a) Example run. (b) Estimation error for ten runs. (c) Magnitude of variance for ten runs.

Fig. 3: Illustration of the performance of the context-aware filter on a non-moving scalar system.



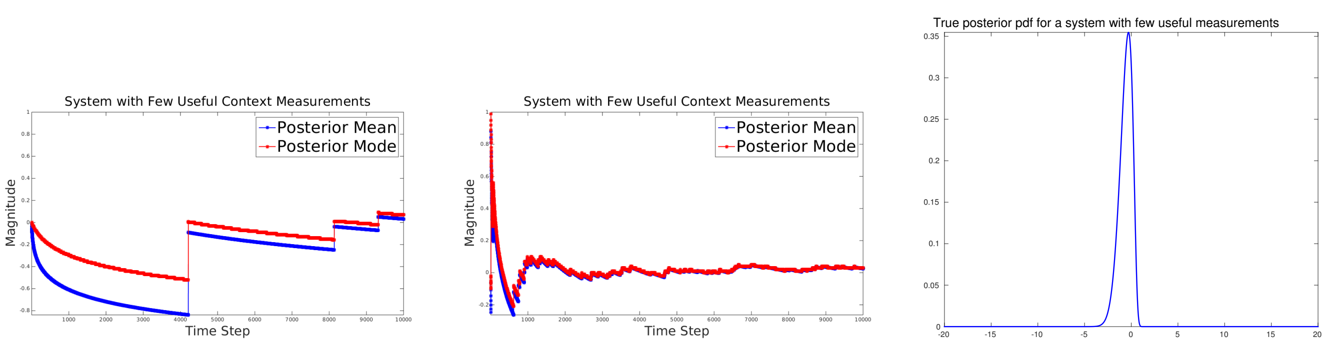
(a) Example run. Note that each axis represents one state of the system.

(b) Total number of context measurements equal to 1 (out of 24) over time for the example run in Figure 4a.



(c) Estimation error for ten runs. (d) Trace of the covariance matrix for ten runs.

Fig. 4: Illustration of the performance of the context-aware filter on an unstable system.



(a) System with $\mu_0 = 0, \sigma_0 = 1, v_k = 1$, $a_k = -3.5$. (b) System with $\mu_0 = 0, \sigma_0 = 1, v_k = 1$, $a_k = -2.5$. (c) True posterior pdf of the system in Figure 5a after 1000 steps (same as in Table I).

Fig. 5: Detailed analysis of two systems from Table I.

C. Shape of the True Posterior Distribution

In this subsection, we provide simulation results in order to inspect the shape of the true posterior distribution and to justify the Gaussian approximation used in the context-aware filter. Since we cannot derive a closed-form expression for the true posterior after more than one update, we simulate multiple different systems, compute the posterior numerically and analyze its properties in order to compare it to the approximating

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

TABLE I: Absolute difference between the true posterior distribution's mean and mode (in one dimension) after 1000 updates.

	$\mu_0 = 0$ $\sigma_0 = 1$	$\mu_0 = 0$ $\sigma_0 = 3$	$\mu_0 = 0$ $\sigma_0 = 5$	$\mu_0 = 0$ $\sigma_0 = 7$	$\mu_0 = 0$ $\sigma_0 = 9$
$v_k = 1$ $a_k = -0.5$	0.0049	0.0051	0.0035	0.0007	0.0017
$v_k = 1$ $a_k = -1.5$	0.0019	0.0018	0.0033	0.0024	0.0071
$v_k = 1$ $a_k = -2.5$	0.0353	0.0144	0.0218	0.0309	0.0193
$v_k = 1$ $a_k = -3.5$	0.3095	0.7392	0.1132	1.3231	1.567
$v_k = 1$ $a_k = -4.5$	0.1986	0.6484	1.0011	1.2933	1.5569

Gaussian. We aim to show that the true posterior is sufficiently symmetric so that the Gaussian approximation is reasonable.

One measure of symmetry is the closeness between the distribution's mean and mode – the more symmetric a distribution is, the closer its mean and mode are. Thus, we measure the difference between the posterior's mean and mode as indication of its symmetry. We simulate multiple non-moving scalar systems with access to one context measurement; in different systems, we vary the value of the offset parameter in the probit function, a_k,[8] and the system's initial covariance. For each system, we record the absolute difference between the posterior's mean and mode after 1000 updates.

Table I presents the results. When a_k is close to the true mean, 0, the difference between the mean and the mode is very small, i.e., the posterior is very symmetric. As a_k gets larger, the difference becomes bigger, which means that the posterior is more skewed. This is due to the fact that for these systems only measurements of -1 are observed during the 1000 simulation steps because the probability of receiving a measurement of 1 is low ($\leq 10^{-4}$). To explore this issue, we simulate two of these systems for longer time; the results are shown in Figure 5. As the number of updates increases, measurements of both 1 and -1 are observed, resulting in the means and modes getting closer. In addition, the true posterior distribution of the system in Figure 5a is plotted after 1000 steps in Figure 5c – although no measurements of 1 have been observed, the distribution appears very similar to a Gaussian. Thus, we conclude that the posterior distribution is close to symmetric for many systems, especially when context measurements have a high probability of being both 1 and -1.

VII. Context-Aware Estimation of Blood Oxygen Content

To evaluate the effectiveness of the context-aware filter, in this section we apply it to the problem of estimating the O_2 content in the blood, one of the most closely monitored variables in operating rooms. The O_2 content has to be maintained within safe ranges; high values could be toxic whereas low values may lead to organ failure. Thus, controlling the O_2 content is one of clinicians' top priorities during surgery.

[8]Since systems are not moving, it is sufficient to only vary a_k in $v_k^T x + a_k$.

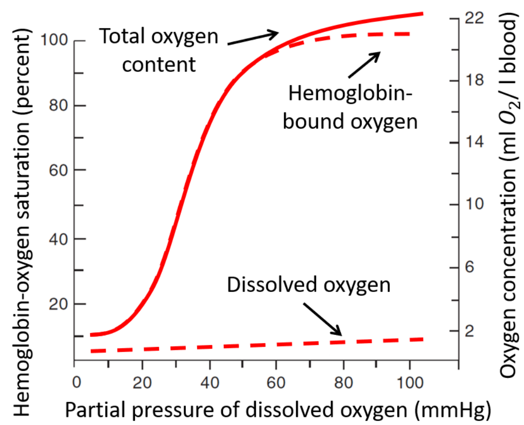


Fig. 6: A typical hemoglobin dissociation curve for O_2. It shows the composition of O_2 content in the blood as well as the shape of the relationship between the overall content and the pressure of dissolved O_2.

Currently, the O_2 content can only be measured through blood gas analysis, which is invasive and not real-time. As a real-time non-invasive alternative, clinicians use a proxy, the hemoglobin-oxygen saturation in the peripheral capillaries (S_pO_2), measured by a pulse oximeter at an extremity (usually a finger tip). S_pO_2 is a good measure of the O_2 content because hemoglobin-bound O_2 accounts for the majority of O_2 in the blood. O_2 appears in two forms in the blood: it is bound to hemoglobin or dissolved in the blood; the relationship between these variables is captured in Figure 6. However, the saturation is usually constant at 100% in healthy people; thus, when reduced S_pO_2 is observed, the O_2 content has already decreased and is potentially entering the steep portion of the curve in Figure 6 where the patient might be in a critical state.

In contrast, estimating the partial pressure of dissolved O_2 (P_aO_2) is proactive because large drops in P_aO_2 are observed before a sharp decrease in the O_2 content, i.e., when the patient is still in the top right portion of the curve in Figure 6. Currently, measuring P_aO_2 also requires blood gas analysis. It is, however, possible to relate other available (real-time and non-invasive) measurements to P_aO_2; in particular, one could use the available pulmonary measurements (e.g., partial pressures of inhaled and exhaled O_2) and construct a (parameterized) model relating the measurements to the state. Once P_aO_2 is estimated, it is also possible to obtain an estimate of the O_2 content by using a (parameterized) functional form of the curve in Figure 6. Thus, the problem addressed in this section is to **estimate both the O_2 content and P_aO_2 using only the non-invasive pulmonary measurements**.

If the model parameters were known, one could use standard filtering techniques to perform the estimation task. However, there are two confounding factors when identifying these parameters from data: 1) the physiological model only captures general trends and does not have great predictive power; 2) the available data is noisy and insufficient to obtain good parameter estimates. Thus, instead of identifying the parameters for each patient, we use population averages for the parameters (as obtained from medical literature) and augment the measurement model with context measurements in order to improve the overall estimation accuracy.

The next subsection provides a summary of the physiological model mapping the measurements to the state (including a general-trends dynamic model for the state). Then we introduce two classes of context measurements as derived from medical device data that is not directly used as a measurement. Finally, we provide the case-study evaluation.

A. Physiological Model

This subsection presents the dynamic physiological model for the O_2 content and for P_aO_2. In the interest of space, only a summary of the model is provided. For a full description of the modeling process, please refer to our preliminary work [2].

At a high level, the circulation of O_2 can be described as follows. As O_2 is inhaled, it reaches the lungs and the alveoli where O_2 enters the blood stream through diffusion. The pulmonary veins carry O_2 to the heart, which pumps O_2 into the arteries and eventually to the peripheral capillaries where metabolism occurs. Metabolism burns O_2 and produces carbon dioxide (CO_2). The CO_2-rich blood is transported via the veins back to the heart, whence it is pumped into the pulmonary arteries that take it to the lungs for a new round of diffusion. A simplified schematic of this process is presented in Figure 7; variables starting with a P denote partial pressures, and variables starting with a C denote concentration (in the blood only); the subscripts denote the corresponding location.

The process of diffusion is complicated by the fact that some blood does not pass through the lungs (e.g., due to blood draining directly into the cavity of the left ventricle through the thebesian veins [50]). Thus, as shown in Figure 8, the shunted blood remains CO_2-rich whereas blood that passes through the lungs diffuses until the partial pressures of O_2 in the blood and the lungs are equal.

By using the intuition from Figures 7 and 8 and two widely used equations from the medical literature, namely the oxygen content equation and the alveolar gas equation [50], we arrive at the final model:

$$\begin{aligned} a_{k+1} &= (1-f)(1.34Hb + 0.003(c_1 u_k + c_{2,k} e_k)) \\ &\quad + f(a_k - \mu) + v_{1,k} \\ e_{k+1} &= e_k + v_{2,k} \\ y_k &= e_k + w_k, \end{aligned} \tag{21}$$

where a_k is the arterial O_2 content, e_k is the partial pressure of exhaled CO_2, Hb is the concentration of hemoglobin in the blood, u_k is the percent of O_2 in inhaled air (as input by clinicians), μ is the effect of metabolism on the O_2 content, f is the proportion of shunted blood, c_1 and $c_{2,k}$ are known constants, and $v_{1,k}$, $v_{2,k}$ and w_k are white Gaussian noises. Finally, y_k denotes the available continuous measurement, the partial pressure of exhaled CO_2 (denoted by $EtCO_2$).

As discussed above, we use population averages for the parameters in (21), μ and Hb; f can be estimated through an initial blood gas measurement [2]. Note that only one of the states in (21) is observed through a continuous measurement. The next subsection describes the context measurements used to estimate the other state, namely the O_2 content.

B. Context Measurements

In order to estimate the O_2 content, we introduce two classes of context measurements as derived from medical device data that is not used directly in (21). The first context measurement can be obtained by using the intuition from Figure 6. Note that as soon as S_pO_2 drops below a certain threshold, the O_2 content is almost entirely determined by hemoglobin-bound O_2. Furthermore, by using the oxygen content equation [50], one can conclude that $C_aO_2 < (1.34 * S_pO_2)Hb$. Thus, we introduce a binary context measurement b_k^1 that is equal to 1 if $S_pO_2 < 99\%$ and -1, otherwise. The parameters in the observation model are set to $v_k^1 = [1 \;\; 0]^T$ and $a_k^1 = -(1.34 * 0.99)Hb$ (once again, a population average is used for Hb).

The second class of context measurements aim to capture the effect of three clinician inputs that are not used in the model directly but do affect the patient's state: the volume of inhaled air (V_t), respiratory rate (RR) and peak inspiratory pressure (PIP). Although mapping these inputs to the O_2 content requires knowledge of multiple non-identifyable parameters (e.g., lung thickness), it is possible to track relative changes in the O_2 content (as caused by relative input changes) once a baseline is established. In particular, we construct a signal s_k that represents the "expected" amount of diffused O_2, up to the unknown parameters (refer to our prior work for the exact functional form of s_k [2]). We initialize s_k (say, at time q) with a single blood gas measurement of the O_2 content, and then track relative changes in s_k, which correspond to relative changes in the O_2 content. Thus, binary context measurements b_k^2, b_k^3, b_k^4, and b_k^5 are introduced that are equal 1 when $s_k < 0.5s_q$, $s_k < 0.8s_q$, $s_k > 1.2s_q$ and $s_k > 1.5s_q$, respectively. The context model parameters are also set accordingly, e.g., $v_k^2 = [1 \;\; 0]^T, a_k^2 = -0.5C_aO_2(q)$, where $C_aO_2(q)$ is the measured O_2 content at time q.

C. Evaluation

To evaluate the filter's performance, we use real-patient data collected during infant lung lobectomies performed at CHOP. A lung lobectomy is the incision of a cyst from the patient's lung; lobectomies often require one-lung ventilation in order to keep the perioperative lung still. In infants, one lung is often not enough to provide sufficient O_2; hence, critical drops in O_2 content are frequently observed during such surgeries.

For evaluation purposes, we use only cases that have at least two blood gas measurements available; the dataset consists of 51 such cases overall. As noted above, the first blood gas measurement is used to initialize s_k; each subsequent blood gas measurement is used as ground truth for evaluation purposes. Finally, the available blood gas data only contain P_aO_2 measurements, hence **only P_aO_2 estimates are evaluated**.

Figure 9 presents the absolute estimation errors of the context-aware filter, with all patient measurements stacked together for easier visualization. We compare the performance of the context-aware filter with another P_aO_2 estimation approach that also requires one blood gas measurement for initialization; we refer to this algorithm as the "F_iO_2-based estimator" (F_iO_2 denotes the fraction of O_2 in inhaled air – it is denoted by u_k in (21)). The context-aware filter's average

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

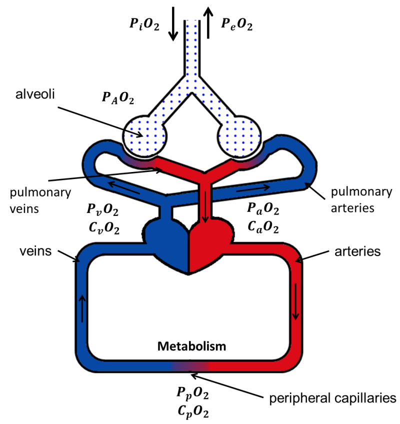


Fig. 7: A simplified schematic model of O_2 variables in the respiratory and cardiovascular systems.

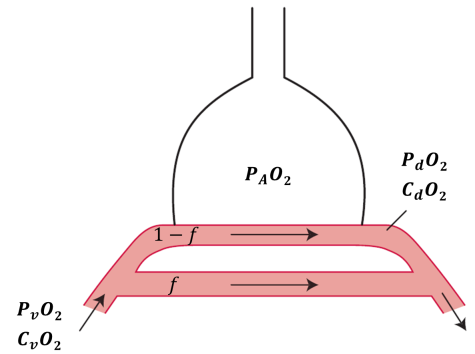


Fig. 8: An illustration of shunted vs. non-shunted blood dynamics in the lung. O_2-rich non-shunted blood participates in diffusion and then mixes with CO_2-rich shunted blood.

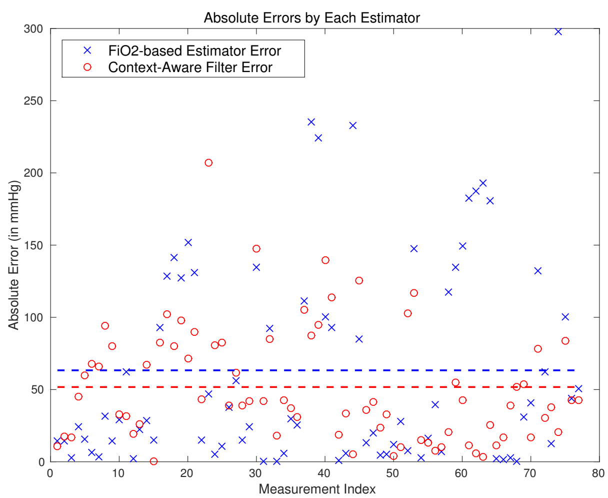


Fig. 9: Absolute errors for each of the two compared P_aO_2 estimators. Red dashed line shows the average error of the context-aware filter, whereas blue dashed line indicates the average error of the F_iO_2-based estimator.

estimation error is about 20% lower (51.7 mmHg vs. 63.3 mmHg). More importantly, the context-aware filter results in much fewer outliers (one error above 150 mmHg as compared to 10 for the F_iO_2-based estimator); this illustrates arguably the biggest benefit of context – providing good information in cases with inaccurate models or insufficient measurements. Note that estimation errors of 100 mmHg (or more) are still significant, and further improvements are required to enable automatically closing the loop; yet, the reasonably uniform distribution of the errors suggests that the context-aware filter is not greatly affected by inter-patient variability and is thus a reasonable choice of estimator, once a more accurate model and more precise measurements are obtained.

To further evaluate the context-aware filter's performance, we present two cases, one with good and one with bad estimation performance, respectively. Figure 10a presents a case where context measurements bring a significant improvement; this is due to the fact that the diffusion signal s_k raises alarms indicating s_k is less than 0.8, but not less than 0.5, of the baseline. Thus, the context-aware filter estimates are around 80% of the initial blood gas measurement, i.e., close to the ground truth. In contrast, the F_iO_2-based estimator is heavily affected by the reduced F_iO_2 and produces large errors.

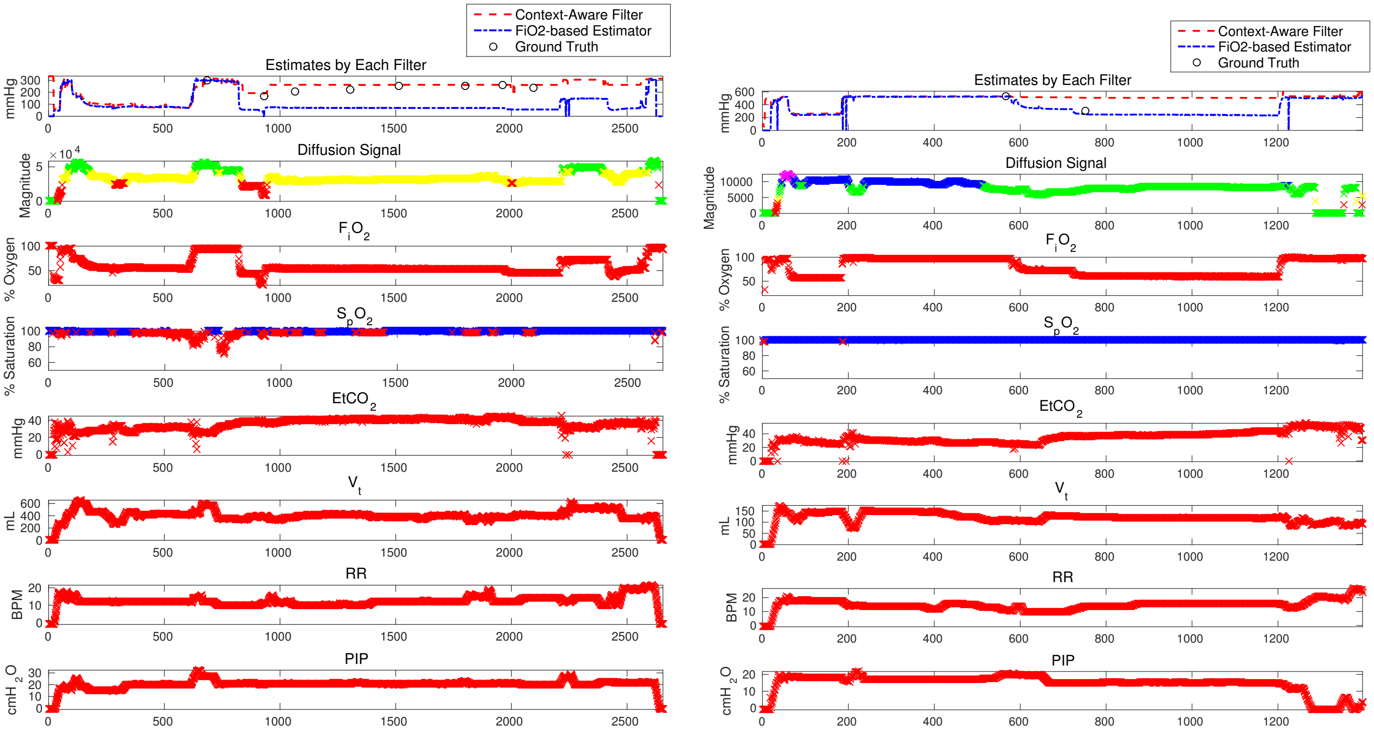
Figure 10b presents an example with a large estimation error by the context-aware filter. In this case, the diffusion signal s_k is too low at the initialization stage, and no low alarms are raised later. A possible explanation for this behavior is a wrong timestamp of the blood gas sample; timestamps are entered manually and are known to be significantly wrong in certain cases [51]. If the baseline had been established around step 420 (which is when clinicians first took action by lowering V_t), low s_k alarms would be raised later, thereby improving the performance of the context-aware filter.

Based on these results, we conclude that the context-aware filter is a promising direction for future work, especially in scenarios with inaccurate models and unobservable states. In addition to improving estimation performance, context greatly reduces worst-case errors, which is critical in a medical setting where good performance for every individual is required.

VIII. Conclusion

This paper addressed the problem of continuous estimation using discrete context measurements. We developed the context-aware filter that approximates the posterior distribution with a Gaussian distribution with the same first two moments. We showed that the filter's expected uncertainty is bounded provided that the probability of receiving context measurements is at least some positive number for all states. Furthermore, we provided an observability-like result that states that the eigenvalues of the filter's covariance matrix converge to zero after repeated updates if and only if a persistence-of-excitation condition is true for the context measurements. In future work, we aim to extend the bounded-uncertainty result to multidimensional systems as well as to analyze conditions under which the filter is asymptotically unbiased.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control



(a) Example case with good estimation by the context-aware filter.

(b) Example case with bad estimation by the context-aware filter.

Fig. 10: Example cases for different scenarios. Red S_pO_2 data points indicate low-S_pO_2 alarms; blue S_pO_2 data points indicate no S_pO_2 alarms. Diffusion signal: red data points indicate $0.5s_q$ alarms; yellow data points indicate $0.8s_q$ alarms; green data points indicate no alarms; blue data points indicate $1.2s_q$ alarms; magenta data points indicate $1.5s_q$ alarms.

APPENDIX A
PROOF OF PROPOSITION 1

First note that the update equation takes the form:

$$p_{k|k}(x) = \frac{p(b_k \mid x)\phi(x;\mu_{k|k-1},\Sigma_{k|k-1})}{\int p(b_k \mid x')\phi(x';\mu_{k|k-1},\Sigma_{k|k-1})dx'} = \frac{\Phi((v_k^T x + a_k)b_k)\phi(x;\mu_{k|k-1},\Sigma_{k|k-1})}{Z_k},$$

where

$$Z_k = \int \Phi((v_k^T x' + a_k)b_k)\phi(x';\mu_{k|k-1},\Sigma_{k|k-1})dx'.$$

The derivation for Z_k is carried out as follows:

$$\begin{aligned} Z_k &= \int \Phi((v_k^T x' + a_k)b_k)\phi(x';\mu_{k|k-1},\Sigma_{k|k-1})dx' \\ &= \mathbb{E}_x\left[\Phi((v_k^T x + a_k)b_k)\right] \\ &= \mathbb{E}_x\left[\mathbb{P}(y \le (v_k^T x + a_k)b_k)\right] = \mathbb{E}_{(x,y)}\left[\mathbb{1}_{y\le(v_k^T x + a_k)b_k}\right] \\ &= \mathbb{P}((v_k^T x + a_k)b_k - y \ge 0) \\ &= \mathbb{P}\left((v_k^T\mu_{k|k-1} + a_k)b_k + z\sqrt{v_k^T\Sigma_{k|k-1}v_k + 1} \ge 0\right) \\ &= \mathbb{P}(z \ge -M_k) = 1 - \Phi(-M_k) = \Phi(M_k), \end{aligned}$$

where y and z are standard Normal random variables independent of each other and of x.

APPENDIX B
PROOF OF PROPOSITION 2

To show that the function $g(x) = \ln(p_{k|k}(x))$ is concave, we need to show that its Hessian (with respect to x) is negative definite. To see this, first note that

$$\begin{aligned} g(x) = &-\ln(Z_k) + \ln(\Phi((v_k^T x + a_k)b_k)) \\ &- \ln(\sqrt{(2\pi)^n|\Sigma_{k|k-1}|}) \\ &- \frac{1}{2}(x-\mu_{k|k-1})^T\Sigma_{k|k-1}^{-1}(x-\mu_{k|k-1}). \end{aligned}$$

The first derivative of $g(x)$ is:

$$g'(x) = v_k b_k \alpha((v_k^T x + a_k)b_k) - \Sigma_{k|k-1}^{-1}(x-\mu_{k|k-1}),$$

where $\alpha(x) = \phi(x;0,1)/\Phi(x)$. The Hessian of $g(x)$ is:

$$\begin{aligned} g''(x) &= v_k v_k^T b_k^2[-\alpha((v_k^T x + a_k)b_k)((v_k^T x + a_k)b_k) \\ &\quad - \alpha^2((v_k^T x + a_k)b_k)] - \Sigma_{k|k-1}^{-1} \\ &= -v_k v_k^T h((v_k^T x + a_k)b_k) - \Sigma_{k|k-1}^{-1}. \end{aligned}$$

Since $v_k v_k^T$ is positive semidefinite and $\Sigma_{k|k-1}$ is positive definite, it remains to show that the term $h((v_k^T x + a_k)b_k)$ is non-negative; but this is true as shown in Proposition 4.

APPENDIX C
PROOF OF PROPOSITION 3

First note that

$$\mu_{k|k} = \int x' \frac{\Phi((v_k^T x' + a_k)b_k)\phi(x';\mu_{k|k-1},\Sigma_{k|k-1})}{Z_k}dx'.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

We compute the mean in closed form, similar to the derivation in Chapter 3.9 in [38], by computing the gradient with respect to $\mu_{k|k-1}$ of the following two equivalent expressions for Z_k:

$$\int \Phi((v_k^T x' + a_k)b_k)\phi(x'; \mu_{k|k-1}, \Sigma_{k|k-1})dx' = \Phi(M_k). \tag{22}$$

The corresponding derivatives are:

$$\begin{aligned}\frac{\partial Z_k}{\partial \mu_{k|k-1}} &= \int \Sigma_{k|k-1}^{-1}(x' - \mu_{k|k-1})\Phi((v_k^T x' + a_k)b_k) \\ &\quad \cdot \phi(x'; \mu_{k|k-1}, \Sigma_{k|k-1})dx' \\ &= b_k v_k \frac{\phi(M_k; 0, 1)}{\sqrt{v_k^T \Sigma_{k|k-1} v_k + 1}},\end{aligned}$$

where we used the fact that $\partial\Phi(x)/\partial x = \phi(x)$. Note that the first term in the integral on the left-hand side is $Z_k \Sigma_{k|k-1}^{-1}\mu_{k|k}$. The second term is $Z_k \Sigma_{k|k-1}^{-1}\mu_{k|k-1}$. Therefore, we get

$$Z_k \Sigma_{k|k-1}^{-1}\mu_{k|k} = Z_k \Sigma_{k|k-1}^{-1}\mu_{k|k-1} + v_k \frac{b_k \phi(M_k; 0, 1)}{\sqrt{v_k^T \Sigma_{k|k-1} v_k + 1}}.$$

Thus, we arrive at

$$\mu_{k|k} = \mu_{k|k-1} + b_k \Sigma_{k|k-1} v_k \frac{\alpha(M_k)}{\sqrt{v_k^T \Sigma_{k|k-1} v_k + 1}},$$

where we used the second expression for Z_k to get α. The final expression for $\mu_{k|k}$ is obtained by solving for χ_k in the equation $\alpha(M_k)(v_k^T \Sigma_{k|k-1} v_k + 1)^{-1/2} = (v_k^T \Sigma_{k|k-1} v_k + \chi_k)^{-1}$.

The expression for the covariance matrix is:

$$\Sigma_{k|k} = \hat{\Sigma}_{k|k} - \mu_{k|k}\mu_{k|k}^T, \tag{23}$$

where

$$\hat{\Sigma}_{k|k} = \int x' x'^T \frac{\Phi((v_k^T x' + a_k)b_k)\phi(x'; \mu_{k|k-1}, \Sigma_{k|k-1})}{Z_k} dx'.$$

$\hat{\Sigma}_{k|k}$ is computed in similar to the mean, by computing the Hessians with respect to $\mu_{k|k-1}$ of both sides of (22):

$$\begin{aligned}&\int \Sigma_{k|k-1}^{-1}(x' - \mu_{k|k-1})(x' - \mu_{k|k-1})^T \Sigma_{k|k-1}^{-1} \\ &\quad \cdot \Phi((v_k^T x' + a_k)b_k)\phi(x'; \mu_{k|k-1}, \Sigma_{k|k-1})dx' \\ &\quad - \int \Sigma_{k|k-1}^{-1}\Phi((v_k^T x' + a_k)b_k)\phi(x'; \mu_{k|k-1}, \Sigma_{k|k-1})dx' \\ &\quad = -b_k v_k v_k^T \frac{\phi(M_k; 0, 1)(v_k^T \mu_{k|k-1} + a_k)}{(v_k^T \Sigma_{k|k-1} v_k + 1)^{3/2}}.\end{aligned}$$

Note that one of the terms in the integral on the left-hand side is $Z_k \Sigma_{k|k-1}^{-1}\hat{\Sigma}_{k|k}\Sigma_{k|k-1}^{-1}$. Therefore, we rearrange terms and divide by Z_k to obtain the following:

$$\begin{aligned}\Sigma_{k|k-1}^{-1}\hat{\Sigma}_{k|k}\Sigma_{k|k-1}^{-1} &= \Sigma_{k|k-1}^{-1} + \Sigma_{k|k-1}^{-1}\mu_{k|k}\mu_{k|k-1}^T\Sigma_{k|k-1}^{-1} \\ &\quad + \Sigma_{k|k-1}^{-1}\mu_{k|k-1}\mu_{k|k}^T\Sigma_{k|k-1}^{-1} \\ &\quad - \Sigma_{k|k-1}^{-1}\mu_{k|k-1}\mu_{k|k-1}^T\Sigma_{k|k-1}^{-1} \\ &\quad - b_k v_k v_k^T \frac{\alpha(M_k)(v_k^T \mu_{k|k-1} + a_k)}{(v_k^T \Sigma_{k|k-1} v_k + 1)^{3/2}}.\end{aligned}$$

Finally, we arrive at the expression for $\hat{\Sigma}_{k|k}$:

$$\begin{aligned}\hat{\Sigma}_{k|k} &= \Sigma_{k|k-1} + \mu_{k|k}\mu_{k|k-1}^T + \mu_{k|k-1}\mu_{k|k}^T - \mu_{k|k-1}\mu_{k|k-1}^T \\ &\quad - b_k \Sigma_{k|k-1} v_k v_k^T \Sigma_{k|k-1} \frac{\alpha(M_k)(v_k^T \mu_{k|k-1} + a_k)}{(v_k^T \Sigma_{k|k-1} v_k + 1)^{3/2}}.\end{aligned}$$

Thus, the covariance matrix can be computed by plugging in the expression for $\hat{\Sigma}_{k|k}$ in (23). To simplify it to the final form shown in the Proposition statement, we first plug in the expression for $\mu_{k|k} - \mu_{k|k-1}$ from (6) and then solve for γ_k.

APPENDIX D
PROOF OF THEOREM 1

Consider the (scalar) modified algebraic Riccati equation (MARE) defined as:

$$g_\beta(x) = axa + q - \beta axv(vxv + 1)^{-1}vxa,$$

where $v = \min_i |v_i|$, i.e., the minimum-in-magnitude of all context weights. Note that if $\beta = 1$, then this becomes the standard algebraic Riccati equation, which converges for any σ_0. On the other hand if $\beta = 0$, the covariance matrix diverges for some σ_0 if a is unstable. We use the MARE to bound the expected value of context-aware filter's variance and give conditions on β for which the expectation is bounded.

We first bound the expected variance of the filter using the MARE. From (10), followed by applying the prediction step, we get (by using the simplified notation $\sigma_k = \sigma_{k|k-1}$):

$$\begin{aligned}\mathbb{E}[\sigma_{k+1}] &= \mathbb{E}[a\sigma_k a + q - \theta_m a\sigma_k v_k(v_k\sigma_k v_k + \gamma_k^m)^{-1}v_k\sigma_k a \\ &\qquad - \theta_p a\sigma_k v_k(v_k\sigma_k v_k + \gamma_k^p)^{-1}v_k\sigma_k a] \\ &\le \mathbb{E}[a\sigma_k a + q - \eta a\sigma_k v(v\sigma_k v + \gamma_k^m)^{-1}v\sigma_k a \\ &\qquad - \eta a\sigma_k v(v\sigma_k v + \gamma_k^p)^{-1}v\sigma_k a] \\ &\le \mathbb{E}[a\sigma_k a + q - \eta a\sigma_k v(v\sigma_k v + \min\{\gamma_k^m, \gamma_k^p\})^{-1}v\sigma_k a] \\ &\le \mathbb{E}[a\sigma_k a + q \\ &\qquad - \eta a\sigma_k v\left(v\sigma_k v + \frac{(1 - h(0))(v\sigma_k v) + 1}{h(0)}\right)^{-1} v\sigma_k a] \\ &= \mathbb{E}[a\sigma_k a + q - \rho a\sigma_k v(v\sigma_k v + 1)^{-1}v\sigma_k a] \\ &= \mathbb{E}[g_\rho(\sigma_k)],\end{aligned}$$

where $\rho = \eta h(0) < 1$, θ_m is the probability of $b_k = -1$ (with resulting γ_k^m); θ_p and γ_k^p are their analogues when $b_k = 1$. The first equality is the expected value of σ_{k+1} for each possible value of b_k. The second inequality uses the fact that both $\theta_p, \theta_m \ge \eta$. In the third inequality we discard one of the two negative terms, keeping the one with smaller γ_k (i.e., the one that results in $M_k < 0$; note that $0 < h(x) < 1$ and $h'(x) < 0$, from Proposition 4). The last inequality is true because $h(x) > h(0)$ for any $x < 0$.

The rest of the proof mimics the proof of Theorem 3 in [12]. Consider the sequence $s_{k+1} = g_\rho(s_k)$, with $s_0 = \sigma_0$. We show that $\mathbb{E}[\sigma_k] \le s_k$ using induction. Note that $\mathbb{E}[\sigma_k] \le s_k$ implies:

$$\mathbb{E}[\sigma_{k+1}] \le \mathbb{E}[g_\rho(\sigma_k)] \le g_\rho(\mathbb{E}[\sigma_k]) \le g_\rho(s_k) = s_{k+1},$$

where the first inequality was shown above, and the second and third inequalities are shown in Lemma 1 in [12]. Furthermore,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

as shown in Theorem 3 in [12], s_k is bounded from above, given that $\rho > \overline{\rho}$ ($\overline{\rho} \in [0,1)$, as shown in [12]), i.e.,

$$\mathbb{E}[\sigma_k] \leq s_k \leq M_{\sigma_0}, \forall k.$$

Appendix E
Proof of Lemma 1

The proof proceeds by induction on k. The base case is shown in (10). For the induction step, we assume that $K < N$ updates result in the form in (14), with matrices Γ_k and V_k replaced by Γ_K and V_K, respectively. Given weights v_{k+K+1}, the next discrete update is

$$\Sigma_{k+K+1} = \Sigma_{k+K} - \Sigma_{k+K} v_{k+K+1} \beta^{-1} v_{k+K+1}^T \Sigma_{k+K} \quad (24)$$

where by induction

$$\Sigma_{k+K} = \Sigma_k - \Sigma_k V_K^T (V_K \Sigma_k V_K^T + \Gamma_K)^{-1} V_K \Sigma_k,$$
$$\beta = v_{k+K+1}^T \Sigma_{k+K} v_{k+K+1} + \gamma_{k+K+1}.$$

By rearranging terms and using the block matrix inversion lemma, Equation (24) can now be written as

$$\Sigma_{k+K+1} = \Sigma_k - \begin{bmatrix} \Sigma_k V_K^T \Sigma_k v_{k+K+1} \end{bmatrix} \cdot \cdot \begin{bmatrix} V_K \Sigma_k V_K^T + \Gamma_K & V_K \Sigma_k v_{k+K+1} \\ v_{k+K+1}^T \Sigma_k V_K^T & v_{k+K+1}^T \Sigma_k v_{k+K+1} + \gamma_{k+K+1} \end{bmatrix}^{-1} \cdot \begin{bmatrix} V_K \Sigma_k \\ v_{k+K+1}^T \Sigma_k \end{bmatrix},$$

i.e.,

$$\Sigma_{k+K+1} = \Sigma_k - \Sigma_k \begin{bmatrix} V_K^T & v_{k+K+1} \end{bmatrix} \cdot \cdot \left[\begin{bmatrix} V_K \\ v_{k+K+1} \end{bmatrix} \Sigma_k \begin{bmatrix} V_K^T & v_{k+K+1}^T \end{bmatrix} + \begin{bmatrix} \Gamma_K & 0 \\ 0 & \gamma_{k+K+1} \end{bmatrix} \right]^{-1} \cdot \begin{bmatrix} V_K \\ v_{k+K+1}^T \end{bmatrix} \Sigma_k,$$

which has the desired form of the Riccati (update) equation.

Appendix F
Proof of Theorem 2

To prove sufficiency (<=), let V be the matrix of persistently exciting v^i, i.e., $V = [v^1, \ldots, v^n]^T$; V is square and invertible. Consider the sequence of times $k_1, k_2, \ldots$, where $k_1 = 1$ and $k_{t+1} = k_t + l_{k_t} + 1$; all v^i in V occur between each pair of k_t and k_{t+1} by construction. Using Lemma 1, it suffices to show that the eigenvalues of the covariance sequence

$$\Sigma_{k_{t+1}} = \Sigma_{k_t} - \Sigma_{k_t} V^T (V \Sigma_{k_t} V^T + \Gamma_{k_t})^{-1} V \Sigma_{k_t} \quad (25)$$

converge to 0 almost surely. Note from (10) that no binary update can increase the eigenvalues of Σ_k, so any updates with weights and offsets not in $\mathcal{P}$ can be ignored as they do not affect the convergence.

Diagonalizing $\Sigma_{k_t} = UDU^T$, we rewrite (25):

$$\Sigma_{k_{t+1}} = U(D - D(D + M\Gamma_{k_t} M^T)^{-1} D) U^T, \quad (26)$$

where $M = U^T V^{-1}$. Thus, we conclude that

$$\Sigma_{k_{t+1}} \preceq U(D - D(D + \delta_{k_t}^{max} I)^{-1} D) U^T, \quad (27)$$

where $\delta_{k_t}^{max}$ is the largest eigenvalue of $M\Gamma_{k_t} M^T$, i.e.,

$$\lambda_{k_{t+1}}^i \leq \lambda_{k_t}^i - \frac{(\lambda_{k_t}^i)^2}{\lambda_{k_t}^i + \delta_{k_t}^{max}}.$$

Therefore, using the second Borel-Cantelli Lemma, $\lambda_{k_t}^i \xrightarrow{a.s.} 0$ as long as the sum of the probabilities of events $\{\delta_k^{max} \geq \delta^*\}_t$ (for some $\delta^* > 0$) is infinite. But $\delta_{k_t}^{max}$ is lower-bounded if $\gamma_{k_t}^{max}$ (the largest γ_k between times k_t and k_{t+1}) is bounded from above. From (11), it can be seen that γ_k is upper-bounded if the function h is bounded from below. But for each k, $M_k < 0$ with probability at least

$$\bar{\delta} := \min_{b_k \in \{1,-1\}, (v^i, a^i) \in \mathcal{P}} \Phi(((v^i)^T x^* + a_i) b_k),$$

where x^* is the true (non-moving) state. Thus, h has a non-zero probability of having negative input, i.e., it is bounded from below by $h(0) = \alpha^2(0)$ (note that $h'(x) < 0$, from Proposition 4). Thus, $\sum_t \mathbb{P}[\delta_{k_t}^{max} \geq \delta^* \mid b_{0:k}] = \infty$ because

$$\mathbb{P}[\delta_{k_t}^{max} \geq \delta^* \mid b_{0:k}] \geq \mathbb{P}[h(M_k) \geq \alpha^2(0) \mid b_{0:k}] \geq \mathbb{P}[M_k < 0 \mid b_{0:k}] \geq \bar{\delta}.$$

To prove necessity (=>), note that if (v_k, a_k) is not persistently exciting, there exists a time K such that the set of context weights v_k for $k > K$ do not span $\mathbb{R}^n$, i.e., the matrix V_K of all such weights is not full rank. We now show this implies that there exists at least one λ_k^i that does not go to 0. Returning to (14), there exists a rotation matrix U such that one eigenvector (call it p) of $\Sigma_k U^T$ is aligned with an eigenvector of $V_K^\perp$, the null space of V_K. Consider the matrix

$$G = U(\Sigma_k - \Sigma_k V_K^T (V_K \Sigma_k V_K^T + \Gamma_k)^{-1} V_K \Sigma_k) U^T.$$

G has the same eigenvalues as Σ_{k+K} but the eigenvalue corresponding to p is also an eigenvalue of Σ_k, i.e., this eigenvalue remains unchanged when V_K is not full rank.

Appendix G
Proof of Theorem 3

First note that applying the matrix inversion lemma to the covariance update in (10), we get:

$$\Omega_{k+1} = (\Sigma_k - \Sigma_k v_{k+1} (v_{k+1}^T \Sigma_k v_{k+1} + \gamma_{k+1})^{-1} v_{k+1}^T \Sigma_k)^{-1} = \Sigma_k^{-1} + v_{k+1} \gamma_{k+1}^{-1} v_{k+1}^T.$$

Therefore,

$$\Omega_{k+1}^s = \Omega_{k+1} - \Omega_k = v_{k+1} \gamma_{k+1}^{-1} v_{k+1}^T.$$

The mean at time $k+1$ is equal to (by using the update in (6)):

$$\mu_{k+1} = \mu_k + \Sigma_k v_{k+1} (v_{k+1}^T \Sigma_k v_{k+1} + \chi_{k+1})^{-1} b_{k+1} = \mu_k + \Sigma_k v_{k+1} N_{k+1}^{-1} b_{k+1},$$

where $N_{k+1} = v_{k+1}^T \Sigma_k v_{k+1} + \chi_{k+1}$. Thus, the information mean of the "site" approximation becomes

$$\omega_{k+1}^s = \Omega_{k+1} \mu_{k+1} - \Omega_k \mu_k = \Omega_{k+1} \mu_k + (I + L_{k+1}) v_{k+1} N_{k+1}^{-1} b_{k+1} - \Omega_k \mu_k = \Omega_{k+1}^s \mu_k + \Omega_k \mu_k + (I + L_{k+1}) v_{k+1} N_{k+1}^{-1} b_{k+1} - \Omega_k \mu_k,$$

where $L_{k+1} = v_{k+1} \gamma_{k+1}^{-1} v_{k+1}^T \Sigma_k$, and we used the inverse-lemma expression for Ω_{k+1}.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

APPENDIX H
PROOF OF COROLLARY 3

As shown in Theorem 2, if v_k is persistently exciting, then all eigenvalues of Σ_k converge to 0. To analyze the convergence of the natural parameters of the "site" approximations, first note that the first two derivatives of ψ are as follows:

$$\psi'_{k+1}(x) = -v_{k+1}\alpha((v_{k+1}^T x + a_{k+1})b_{k+1})b_{k+1} \tag{28}$$

$$\psi''_{k+1}(x) = v_{k+1}v_{k+1}^T h((v_{k+1}^T x + a_{k+1})b_{k+1}). \tag{29}$$

We first show that $\Omega_{k+1}^s = v_{k+1}\gamma_{k+1}^{-1}v_{k+1}^T$ converges to $\psi''_{k+1}(\mu_k)$, i.e., that γ_{k+1}^{-1} converges to $h((v_{k+1}^T\mu_k + a_{k+1})b_{k+1})$. But this is clear from (11): as the eigenvalues of Σ_k converge to 0, γ_{k+1}^{-1} converges to $h(M_{k+1})$, and M_{k+1} converges to $(v_{k+1}^T\mu_k + a_{k+1})b_{k+1}$.

As derived in (18), the information mean is $\omega_{k+1}^s = \Omega_{k+1}^s\mu_k + (I + L_{k+1})v_{k+1}N_{k+1}^{-1}b_{k+1}$. First note that N_{k+1}^{-1} converges to $1/\chi_{k+1}$, which in turn converges to $\alpha((v_{k+1}^T\mu_k + a_{k+1})b_{k+1})$, as can be seen from (7). Thus, in order to show that the second term of ω_{k+1}^s converges to $-\psi'_{k+1}(\mu_k)$, it suffices to show that L_{k+1} converges to 0. But this is clear from the definition of L_{k+1} in Theorem 3.

REFERENCES

[1] R. Ivanov, N. Atanasov, M. Pajic, G. Pappas, and I. Lee, "Robust estimation using context-aware filtering," in *53rd Annual Allerton Conference on Communication, Control, and Computing*, 2015, pp. 590–597.

[2] R. Ivanov, N. Atanasov, J. Weimer, M. Pajic, A. Simpao, M. Rehman, G. J. Pappas, and I. Lee, "Estimation of blood oxygen content using context-aware filtering," in *Proceedings of the 7th International Conference on Cyber-Physical Systems*. IEEE Press, 2016, p. 28.

[3] T. Kos, I. Markezic, and J. Pokrajcic, "Effects of multipath reception on GPS positioning performance," in *ELMAR, 2010 PROCEEDINGS*. IEEE, 2010, pp. 399–402.

[4] R. Ivanov, J. Weimer, A. Simpao, M. Rehman, and I. Lee, "Early detection of critical pulmonary shunts in infants," in *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*, ser. ICCPS '15. ACM, 2015, pp. 110–119.

[5] D. Shepard, J. Bhatti, and T. Humphreys, "Drone hack," *GPS World*, vol. 23, no. 8, pp. 30–33, 2012.

[6] S. Peterson and P. Faramarzi, "Iran hijacked US drone, says Iranian engineer," *Christian Science Monitor, December*, vol. 15, 2011.

[7] K. Sun, K. Saulnier, N. Atanasov, G. J. Pappas, and V. Kumar, "Dense 3-d mapping with spatial correlation via gaussian filtering," submitted.

[8] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House, Inc., 2007.

[9] L. Y. Wang, J. Zhang, and G. G. Yin, "System identification using binary sensors," *IEEE Transactions on Automatic Control*, vol. 48, no. 11, pp. 1892–1907, 2003.

[10] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, "Soi-kf: Distributed kalman filtering with low-cost communications using the sign of innovations," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4782–4795, 2006.

[11] M. Fu and C. E. de Souza, "State estimation for linear discrete-time systems using quantized measurements," *Automatica*, vol. 45, no. 12, pp. 2937–2945, 2009.

[12] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1453–1464, 2004.

[13] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[14] T. P. Minka, "Expectation propagation for approximate bayesian inference," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2001, pp. 362–369.

[15] M. F. Huber and U. D. Hanebeck, "The hybrid density filter for nonlinear estimation based on hybrid conditional density approximation," in *Information Fusion, 10th International Conference on*, 2007, pp. 1–8.

[16] M. Milanese and C. Novara, "Set Membership identification of nonlinear systems," *Automatica*, vol. 40, no. 6, pp. 957–975, 2004.

[17] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, 2000.

[18] L. Shi, M. Epstein, and R. M. Murray, "Kalman filtering over a packet-dropping network: A probabilistic perspective," *IEEE Transactions on Automatic Control*, vol. 55, no. 3, pp. 594–604, 2010.

[19] V. Gupta, B. Hassibi, and R. M. Murray, "Optimal lqg control across packet-dropping links," *Systems & Control Letters*, vol. 56, no. 6, pp. 439–446, 2007.

[20] O. C. Imer, S. Yüksel, and T. Başar, "Optimal control of lti systems over unreliable communication links," *Automatica*, vol. 42, no. 9, pp. 1429–1439, 2006.

[21] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *Decision and Control, 46th IEEE Conference on*, 2007, pp. 5492–5498.

[22] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.

[23] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *Transactions on Signal Processing*, vol. 57, no. 2, pp. 451–462, 2009.

[24] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," *Automatica*, vol. 48, no. 10, pp. 2482–2493, 2012.

[25] J. Williams, "Information theoretic sensor management," Ph.D. dissertation, MIT, 2007.

[26] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in sensor networks," in *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*. ACM, 1999, pp. 263–270.

[27] S. MartíNez and F. Bullo, "Optimal sensor placement and motion coordination for target tracking," *Automatica*, vol. 42, no. 4, pp. 661–668, 2006.

[28] I. Hwang, H. Balakrishnan, and C. Tomlin, "State estimation for hybrid systems: applications to aircraft tracking," *IEE Proceedings-Control Theory and Applications*, vol. 153, no. 5, pp. 556–566, 2006.

[29] O. L. V. Costa and S. Guerra, "Stationary filter for linear minimum mean square error estimator of discrete-time markovian jump systems," *Transactions on Automatic Control*, vol. 47, no. 8, pp. 1351–1356, 2002.

[30] S. C. Smith and P. Seiler, "Estimation with lossy measurements: jump estimators for jump systems," *IEEE Transactions on Automatic Control*, vol. 48, no. 12, pp. 2163–2171, 2003.

[31] M. C. F. Donkers, W. P. M. H. Heemels, N. Van de Wouw, and L. Hetel, "Stability analysis of networked control systems using a switched linear systems approach," *IEEE Transactions on Automatic control*, vol. 56, no. 9, pp. 2101–2115, 2011.

[32] S. Paoletti, A. L. Juloski, G. Ferrari-Trecate, and R. Vidal, "Identification of hybrid systems: A tutorial," *European journal of control*, vol. 13, no. 2-3, pp. 242–260, 2007.

[33] Z. Wang, D. W. C. Ho, and X. Liu, "Variance-constrained filtering for uncertain stochastic systems with missing measurements," *IEEE Transactions on Automatic control*, vol. 48, no. 7, pp. 1254–1258, 2003.

[34] M. Blaschko and C. Lampert, "Object localization with global and local context kernels," in *British Machine Vision Conference*, 2009.

[35] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. Fernandez-Madrigal, and J. Gonzalez, "Multi-hierarchical semantic maps for mobile robotics," in *Int. Conf. on Intelligent Robots and Systems*, 2005.

[36] R. Anati, D. Scaramuzza, K. Derpanis, and K. Daniilidis, "Robot localization using soft object detection," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2012, pp. 4992–4999.

[37] N. Atanasov, M. Zhu, K. Daniilidis, and G. Pappas, "Semantic localization via the matrix permanent," in *Robotics: Science and Systems*, 2014.

[38] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[39] H. Nickisch and C. Rasmussen, "Approximations for binary gaussian process classification," *Journal of Machine Learning Research (JMLR)*, vol. 9, no. Oct, pp. 2035–2078, 2008.

[40] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT press, 2005.

[41] L. He, Y. Qi, and J. Chen, "Optimal estimation algorithm design under event-based sensor data scheduling," in *In Proceedings of the 6th IFAC Workshop on Distributed Estimation and Control in Networked Systems (NecSys16)*, 2016, pp. 157–162.

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2797839, IEEE Transactions on Automatic Control

[43] S. Dharmadhikari and K. Joag-Dev, *Unimodality, convexity, and applications*. Elsevier, 1988.

[44] M. R. Sampford, "Some inequalities on mill's ratio and related functions," *The Annals of Mathematical Statistics*, vol. 24, no. 1, pp. 130–132, 1953.

[45] M. Green and J. B. Moore, "Persistence of excitation in linear systems," *Systems & Control Letters*, vol. 7, no. 5, pp. 351–360, 1986.

[46] G. Dehaene and S. Barthelmé, "Bounding errors of expectation-propagation," in *Advances in Neural Information Processing Systems*, 2015, pp. 244–252.

[47] ——, "Expectation propagation in the large-data limit," *arXiv preprint arXiv:1503.08060*, 2015.

[48] N. Atanasov, R. Tron, V. Preciado, and G. Pappas, "Joint Estimation and Localization in Sensor Networks," in *IEEE Conf. on Decision and Control*, 2014, pp. 6875–6882. [Online]. Available: http://dx.doi.org/10.1109/CDC.2014.7040469

[49] P. J. Bickel and K. A. Doksum, *Mathematical statistics: basic ideas and selected topics*. CRC Press, 2015, vol. 2.

[50] J. B. West, *Respiratory Physiology: The Essentials*. Lippincott Williams & Wilkins, 2012.

[51] A. F. Simpao, E. Y. Pruitt, S. D. Cook-Sather, H. Gurnaney, and M. Rehman, "The reliability of manual reporting of clinical events in an anesthesia information management system (AIMS)," *Journal of Clinical Monitoring and Computing*, vol. 26, no. 6, pp. 437–439, 2012.

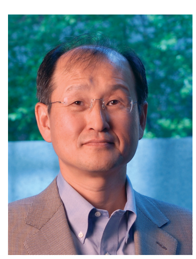
Radoslav Ivanov (S'15) received the B.A. degree in computer science and economics from Colgate University, Hamilton, NY, USA, in 2011, and the Ph.D. degree in computer and information science from the University of Pennsylvania, Philadelphia, PA, USA, in 2017. He is currently a postdoctoral fellow at the University of Pennsylvania. His research interests include the design and control of cyber-physical systems (CPS), in particular, automotive and medical CPS, and predictive and retrospective analysis of medical patient data.

Nikolay Atanasov received the Ph.D. degree in electrical and systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 2015. He is currently an Assistant Professor at the Department of Electrical and Computer Engineering, University of California, San Diego, CA, USA. His research focuses on robotics, control theory, and machine learning and in particular on estimation and control techniques that increase the autonomy, reliability, efficiency, and versatility of robotic sensing systems in applications such as environmental monitoring, localization and mapping, security and surveillance. Dr. Atanasov's work has been recognized by the best Ph.D. dissertation award in electrical and systems engineering at the University of Pennsylvania in 2015 and as the best conference paper award finalist at ICRA, 2017.

Miroslav Pajic (S'06-M'13) received the Dipl. Ing. and M.S. degrees in electrical engineering from the University of Belgrade, Serbia, in 2003 and 2007, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Pennsylvania, Philadelphia, in 2010 and 2012, respectively. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at Duke University. He also holds a secondary appointment in the Computer Science Department. His research interests focus on the design and analysis of cyber-physical systems and in particular distributed/networked control systems, embedded systems, and high-confidence control system. Dr. Pajic received various awards including the National Science Foundation (NSF) CAREER award, Office of Naval Research (ONR) Young Investigator Program Award, 2011 ACM SIGBED Frank Anger Memorial Award, the Joseph and Rosaline Wolf Best Dissertation Award from Penn, and numerous best paper awards.

James Weimer (M'06) received the B.S. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 2005, and the M.S. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2007 and 2010, respectively. He is currently a Research Assistant Professor of Computer and Information Science at the University of Pennsylvania. His research interests include the design and analysis of cyber-physical systems with application to medical devices/monitors, networked systems, building energy management, and security.

George J. Pappas received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 1998. He is currently the Joseph Moore Professor and Chair of the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. He also holds a secondary appointment with the Department of Computer and Information Sciences and the Department of Mechanical Engineering and Applied Mechanics. He is a Member of the GRASP Lab and the PRECISE Center. He had previously served as the Deputy Dean for Research with the School of Engineering and Applied Science. His research interests include control theory and, in particular, hybrid systems, embedded systems, cyber-physical systems, and hierarchical and distributed control systems, with applications to unmanned aerial vehicles, distributed robotics, green buildings, and biomolecular networks. Dr. Pappas has received various awards, such as the Antonio Ruberti Young Researcher Prize, the George S. Axelby Award, the Hugo Schuck Best Paper Award, the George H. Heilmeier Award, the National Science Foundation PECASE award and numerous best student papers awards.



Insup Lee (S'80-M'83-SM'97-F'01) is the Cecilia Fitler Moore Professor in the Department of Computer and Information Science and Director of PRECISE Center, which he co-founded in 2008, at the University of Pennsylvania. He also holds a secondary appointment in the Department of Electrical and Systems Engineering. His research interests include cyber-physical systems (CPS), real-time systems, embedded systems, high-confidence medical device systems, formal methods and tools, run-time verification, software certification, and trust management. The theme of his research activities has been to assure and improve the correctness, safety, and timeliness of life-critical embedded systems. He is a fellow of IEEE and received IEEE TC-RTS Outstanding Technical Achievement and Leadership Award in 2008.