Filtering Over Non-Gaussian Channels: The Role of Anytime Capacity

Zhenyu Liu, *Graduate Student Member, IEEE*, Andrea Conti, *Fellow, IEEE*, Sanjoy K. Mitter, *Life Fellow, IEEE*, and Moe Z. Win, *Fellow, IEEE*

***Abstract*—Filtering over noisy channels is of interest in many network applications, in which a node infers a time-varying state by using messages received from another node that can observe such a state. This letter explores filtering with general models for state disturbance and communication channels by deriving a sufficient condition for which the estimation error is bounded. Specifically, the sufficient condition is expressed in terms of anytime capacity, a notion that characterizes the maximum sequential communication rate. The joint design of encoder and estimator with bounded estimation error is also presented.**

***Index Terms*—Non-Gaussian filtering, networks, anytime capacity, distributed inference.**

I. Introduction

Filtering [1]–[3] is a key enabler for many applications such as localization and navigation, Internet-of-Things, and target tracking [4]–[9]. In many networked systems, nodes that observe the unknown states and nodes that aim to estimate the states are not co-located. In such systems, the nodes observing the unknown states transmit messages to other nodes via inter-node communications, and the received messages are used for estimating the unknown states. To achieve desirable estimation accuracy, the transmitted messages need to be carefully encoded to convey useful information for filtering and to be robust in the presence of channel noise and interference.

This letter investigates the problem of estimating an evolving state using messages received via a noisy channel. As the state is time-varying, both the encoding at the transmitter and the estimation at the receiver need to be performed in real time. Consequently, conventional channel coding techniques [10] cannot be directly applied for generating the transmitted messages, as these techniques can incur high latencies. Moreover, efficient usage of the communication channel requires the consideration of nonlinear encoding strategies, which are more difficult to design and analyze compared to linear ones.

Coding and estimation techniques have been studied in the context of communication, control, and inference [11]–[24]. In particular, a coder-controller was designed in [16] for stabilizing a linear system based on data with limited rates. The notions of anytime reliability and anytime capacity were introduced in [17], where these notions were used to derive necessary and sufficient conditions for stabilizing systems over noisy channels. In particular, anytime capacity characterizes the maximum possible rate at which data can be transmitted reliably in a sequential communication system with encoding and decoding performed in real time. Existing works on estimation using received messages typically consider specific types of channels, such as noiseless channels and Gaussian channels, and hence techniques for more general channels remain to be designed.

This letter aims to derive conditions under which desirable accuracy is achieved in a filtering over noisy channels problem, where a general system model is considered without assuming that the state disturbance and communication noise are Gaussian. Key contributions of this letter are as follows:

- we establish a sufficient condition for the channel in terms of its anytime capacity under which the estimation error at the receiver node is bounded over time; and
- we design a real-time encoder and estimator at the transmitter and receiver, respectively, and prove that the designed estimator has bounded error over time if the sufficient condition is satisfied.

Notations: Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. The expectation of x is denoted by $\mathbb{E}\{\mathsf{x}\}$. For non-negative integers $s \leqslant t$, notation $\mathsf{x}_{s:t}$ represents the concatenation of $\mathsf{x}_s, \mathsf{x}_{s+1}, \ldots, \mathsf{x}_t$, whereas $\{s:t\}$ represents $\{s, s+1, \ldots, t\}$. The sets of real numbers and non-negative integers are denoted by $\mathbb{R}$ and $\mathbb{N}$, respectively. The largest integer less than or equal to $x \in \mathbb{R}$ is denoted by $\lfloor x \rfloor$. Logarithms of $x > 0$ with base 2 and base $\mu > 0$ are denoted by $\log x$ and $\log_\mu x$, respectively. The Euclidean norm and the ith entry of $\boldsymbol{x}$ are denoted by $\|\boldsymbol{x}\|$ and $[\boldsymbol{x}]_i$, respectively. The transpose, determinant, and spectral norm (i.e., the largest singular value) of a matrix $\boldsymbol{A}$ are denoted by $\boldsymbol{A}^{\mathrm{T}}$, $\det(\boldsymbol{A})$, and $\|\boldsymbol{A}\|$, respectively.

Manuscript received 21 March 2022; revised 31 May 2022; accepted 23 June 2022. Date of publication 11 July 2022; date of current version 22 July 2022. This work was supported in part by the Office of Naval Research under Grant N00014-16-1-2141 and Grant N62909-22-1-2009, and in part by the Army Research Office through the MIT Institute for Soldier Nanotechnologies under Contract W911NF-13-D-0001. Recommended by Senior Editor V. Ugrinovskii. *(Corresponding author: Moe Z. Win.)*

Zhenyu Liu is with the Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Andrea Conti is with the Department of Engineering and CNIT, University of Ferrara, 44122 Ferrara, Italy.

Sanjoy K. Mitter and Moe Z. Win are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: moewin@mit.edu).

Digital Object Identifier 10.1109/LCSYS.2022.3189972

II. PROBLEM FORMULATION

Consider a linear system with its state at time t denoted by an n-dimensional random vector $\boldsymbol{\theta}_t$. In particular, $\boldsymbol{\theta}_t$ satisfies

$$\boldsymbol{\theta}_t = \boldsymbol{F}\boldsymbol{\theta}_{t-1} + \mathbf{v}_t \tag{1}$$

where $\boldsymbol{F} \in \mathbb{R}^{n\times n}$ is a real square matrix such that all its eigenvalues have magnitudes no smaller than 1, and $\mathbf{v}_t$ is the state disturbance. A transmitter sends an encoded message $\mathbf{m}_t$ computed based on $\boldsymbol{\theta}_{0:t}$ over a memoryless channel to a receiver at each time t. Let $\mathbf{r}_t$ represent the received message at time t. The receiver evaluates an estimator $\hat{\boldsymbol{\theta}}_t$ of $\boldsymbol{\theta}_t$ based on messages $\mathbf{r}_{0:t}$ received up to time t. In other words, $\hat{\boldsymbol{\theta}}_t$ is a function of $\mathbf{r}_{0:t}$. This letter establishes conditions under which the bth moment of the estimation error is bounded for any $b \geqslant 2$, i.e.,

$$\sup_{t\geqslant 0} \mathbb{E}\left\{\left\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t\right\|^b\right\} < \infty\,. \tag{2}$$

In the special case where $b = 2$, inequality (2) states that the mean-square error (MSE) of $\hat{\boldsymbol{\theta}}_t$ is bounded over time. We consider general models for the state disturbance and communication channel without specifying the distribution of $\mathbf{v}_t$ or the conditional distribution of $\mathbf{r}_t$ given $\mathbf{m}_t$. Instead, we only make a mild assumption that the disturbance satisfies

$$\sup_{t\geqslant 0} \mathbb{E}\left\{\left\|\mathbf{v}_t\right\|^a\right\} < \infty \tag{3}$$

for some $a > b$. This holds if the tail of the distribution for each entry of $\mathbf{v}_t$ is not heavy (e.g., each entry of $\mathbf{v}_t$ is a sub-Gaussian or sub-exponential random variable [25]).

The filtering problem studied here differs from classical ones. Classical filtering techniques, such as Kalman filtering and particle filtering, aim to estimate the unknown states using a set of sensor measurements. These measurements are directly available to the estimator without any loss so that communication is not involved or can be considered ideal. By contrast, we consider a filtering problem where the estimation is performed using messages received via a noisy communication channel with limited capacity. In this problem, the transmitter needs to perform encoding for the transmitted messages in order to protect them against the degradation due to the channel under the capacity limitations. As a result, both the estimation algorithm and the encoding strategy are critical to the filtering accuracy and need to be carefully designed.

Our method is based on the notions of anytime reliability and anytime capacity, introduced in [17] and explained in the rest of this section. Consider a communication system aiming to transmit a sequence of data symbols via a channel. Different from classical block-coding [10] where all the data symbols are available to the encoder beforehand, the encoder of this communication system obtains data symbols sequentially: at each time t, a new data symbol $\mathbf{s}_t$ from an alphabet with 2^r elements is available to the encoder. Based on available data symbols $\mathbf{s}_{0:t}$, the encoder generates a message $\mathbf{m}_t$ and transmits it at each time t. In other words, the encoder performs real-time encoding without waiting for the entire data symbol sequence to be available. Such an encoder is called an anytime encoder and its data rate is r bits per channel use.

The received message corresponding to the transmitted $\mathbf{m}_t$ is denoted by $\mathbf{r}_t$, which can be different than $\mathbf{m}_t$ due to impairments in the channel. At each time, the receiver decodes all the symbols that have been sent by the transmitter. Specifically, at time t, the receiver generates an estimator $\hat{\mathbf{s}}_{t'}(\mathbf{r}_{0:t})$ of $\mathbf{s}_{t'}$ based on $\mathbf{r}_{0:t}$ for all $t' \in \{0:t\}$. In other words, the receiver performs decoding without waiting for the transmitter to complete sending the entire data symbol sequence. Such a receiver is called an anytime decoder. One method for constructing anytime encoder and decoder is by using the connection between communications and control over noisy channels. An example architecture of anytime encoder and decoder can be found in [17].

The system described above is referred to as a rate r sequential communication system. This system achieves *anytime reliability* α if there exists a constant K such that

$$\mathbb{P}\left\{\hat{\mathbf{s}}_{0:t'}(\mathbf{r}_{0:t}) \neq \mathbf{s}_{0:t'}\right\} \leqslant K2^{-\alpha(t-t')}, \qquad \forall t' \in \{0:t\} \tag{4}$$

where $\hat{\mathbf{s}}_{t_1:t_2}(\mathbf{r}_{0:t})$ represents the concatenation of estimators $\hat{\mathbf{s}}_{t_1}(\mathbf{r}_{0:t}), \hat{\mathbf{s}}_{t_1+1}(\mathbf{r}_{0:t}), \ldots, \hat{\mathbf{s}}_{t_2}(\mathbf{r}_{0:t})$ for $0 \leqslant t_1 \leqslant t_2 \leqslant t$. The α*-anytime capacity* $\breve{C}(\alpha)$ of the channel is defined as the smallest upper bound of r such that a rate r sequential communication system that achieves anytime reliability α exists. The anytime capacity of a channel is no larger than its Shannon capacity. We refer to [17], [26] for other properties of anytime capacity.

III. SUFFICIENT CONDITION FOR BOUNDEDNESS OF ESTIMATION ERROR

We now provide a sufficient condition for (2): if the anytime capacity of the channel is above a threshold determined by $\boldsymbol{F}$, then there exist a transmitter and a receiver such that (2) holds.

Proposition 1: Let $\rho(\boldsymbol{F})$ represent the spectral radius, i.e., the largest of the magnitudes of all the eigenvalues, of $\boldsymbol{F}$. For any $2 \leqslant b < a$, if there exists an α such that

$$\alpha > \frac{ab}{a-b}\rho(\boldsymbol{F}) \tag{5}$$

and the α-anytime capacity $\breve{C}(\alpha)$ of the channel satisfies

$$\breve{C}(\alpha) > \log\big(|\det(\boldsymbol{F})|\big) \tag{6}$$

then there exist a transmitter and a receiver such that (2) holds.

Proof: The proposition is proved by constructing a transmitter as well as a receiver and showing that they achieve (2). Here, we show the architecture of the transmitter and the receiver, whereas a sketch for the proof of (2) is presented in Appendix B.

A block diagram for the transmitter and the receiver is shown in Fig. 1, with each block explained as follows.

Transmitter blocks:

(a) Linear transformation via $\boldsymbol{M}^{-1}$*:* A linear transformation of the state $\boldsymbol{\theta}_t$ is performed to obtain $\boldsymbol{\psi}_t$ given by

$$\boldsymbol{\psi}_t := \boldsymbol{M}^{-1}\boldsymbol{\theta}_t\,. \tag{7}$$

Here, $\boldsymbol{M}$ is a matrix that transforms $\boldsymbol{F}$ to its real Jordan canonical form [27, Ch. 3] $\boldsymbol{J}$ as

$$\boldsymbol{F} = \boldsymbol{M}\boldsymbol{J}\boldsymbol{M}^{-1}.$$

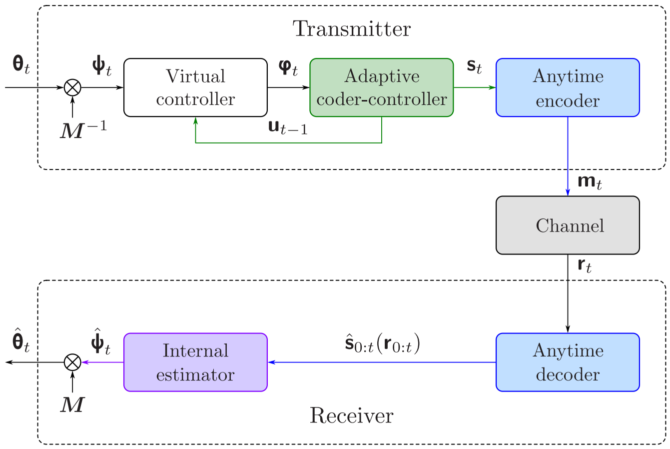

Fig. 1. Block diagram for the designed transmitter and receiver.

(b) Virtual controller[1]*:* An auxiliary controlled state $\boldsymbol{\varphi}_t$ is computed as

$$\boldsymbol{\varphi}_t := \boldsymbol{\psi}_t + \sum_{t'=0}^{t-1} \boldsymbol{J}^{t-1-t'} \mathbf{u}_{t'} \tag{8}$$

where $\mathbf{u}_{t'}$ is a control signal generated by the adaptive coder-controller at time t'. Combining (1), (7), and (8) gives

$$\boldsymbol{\varphi}_t = \boldsymbol{J}\boldsymbol{\varphi}_{t-1} + \mathbf{u}_{t-1} + \boldsymbol{M}^{-1}\mathbf{v}_t\,.$$

(c) Adaptive coder-controller: Reference [16, Th. 2.1] can be extended to show that there exist a coder and a controller generating control $\mathbf{u}_t$ at time t such that the bth moment of $\|\boldsymbol{\varphi}_t\|$ is bounded over time, i.e.,

$$\sup_{t\geqslant 0} \mathbb{E}\big\{\|\boldsymbol{\varphi}_t\|^b\big\} < \infty\,. \tag{9}$$

We refer to this pair of coder and controller, introduced in [16], as adaptive coder-controller as it employs adaptive quantization. The coder in this pair is referred to as an adaptive coder. Specifically, the adaptive coder performs vector quantization and generates a classification vector $\mathbf{w}_j$ at time $j\tau$ for all $j \in \mathbb{N}$, where τ is the quantization period parameter for the coder. The classification vector $\mathbf{w}_j$ is given by

$$\mathbf{w}_j := \boldsymbol{\omega}\big((\boldsymbol{\varphi}_{j\tau} - \hat{\boldsymbol{\varphi}}_{\mathrm{q},j\tau})/\mathsf{l}_{j\tau}\big) \tag{10}$$

where the random vector $\hat{\boldsymbol{\varphi}}_{\mathrm{q},j\tau}$ is an estimator of $\boldsymbol{\varphi}_{j\tau}$ based on previous classification vectors $\mathbf{w}_{0:j-1}$, the random variable $\mathsf{l}_{j\tau}$ is an adaptive scaling factor, and $\boldsymbol{\omega}(\cdot)$ represents the vector quantization operator described in Appendix A. Equation (10) can be interpreted as follows: at time $j\tau$, the error $\boldsymbol{\varphi}_{j\tau} - \hat{\boldsymbol{\varphi}}_{\mathrm{q},j\tau}$ of the estimator $\hat{\boldsymbol{\varphi}}_{\mathrm{q},j\tau}$ based on $\mathbf{w}_{0:j-1}$ is scaled by $\mathsf{l}_{j\tau}$, and quantization is performed on such scaled error to generate the new classification vector $\mathbf{w}_j$. Expressions of $\hat{\boldsymbol{\varphi}}_{\mathrm{q},t}$ and l_t for an arbitrary time $t \in \mathbb{N}$ are given in (20) of Appendix A.

The classification vector $\mathbf{w}_j$ belongs to a finite set and thus can be converted to data symbols via a bijection. Specifically,

[1]The aim of this block is to compute an auxiliary controlled state for the encoding procedure. There is no physical controller in the block and this is the reason for the word "virtual."

Algorithm 1 Signal Processing by the Transmitter at Time t

Input: State $\boldsymbol{\theta}_t$
Output: Transmitted message $\mathbf{m}_t$
1: Perform linear transformation $\boldsymbol{\psi}_t \leftarrow \boldsymbol{M}^{-1}\boldsymbol{\theta}_t$
2: Compute auxiliary controlled state $\boldsymbol{\varphi}_t$ according to (8)
3: Compute $\mathbf{w}_j$ according to (10) if $t = j\tau$ for $j \in \mathbb{N}$
4: Compute $\hat{\boldsymbol{\varphi}}_{\mathrm{q},t}$ and l_t according to (20)
5: Compute control signal $\mathbf{u}_t$ according to (21)
6: Evaluate symbol $\mathbf{s}_t$ by applying $\boldsymbol{f}_{\mathrm{s}}(\cdot)$
7: Generate $\mathbf{m}_t$ based on $\mathbf{s}_{0:t}$ using the anytime encoder

Algorithm 2 Signal Processing by the Receiver at Time t

Input: Received messages $\mathbf{r}_{0:t}$
Output: Estimator $\hat{\boldsymbol{\theta}}_t$
1: Evaluate $\hat{\mathbf{s}}_{t'}(\mathbf{r}_{0:t})$ for $t' \in \{0:t\}$ using the anytime decoder
2: Evaluate estimator $\hat{\mathbf{w}}_j(\mathbf{r}_{0:t})$ of classification vector $\mathbf{w}_j$ for $j \in \{0:\lfloor (t+1)/\tau \rfloor - 1\}$ by applying $\boldsymbol{f}_{\mathrm{s}}^{-1}(\cdot)$
3: Compute $\hat{\boldsymbol{\varphi}}_{t'}(\mathbf{r}_{0:t})$ and $\hat{\mathsf{l}}_{t'}(\mathbf{r}_{0:t})$ according to (22) for $t' \in \{0:t-1\}$
4: Compute $\hat{\boldsymbol{\psi}}_t$ according to (14)
5: Perform linear transformation $\hat{\boldsymbol{\theta}}_t \leftarrow \boldsymbol{M}\hat{\boldsymbol{\psi}}_t$

the adaptive coder generates symbols $\mathbf{s}_{j\tau\,:(j+1)\tau-1}$ by applying a bijective function $\boldsymbol{f}_{\mathrm{s}}(\cdot)$ on $\mathbf{w}_j$ as

$$\mathbf{s}_{j\tau\,:(j+1)\tau-1} = \boldsymbol{f}_{\mathrm{s}}(\mathbf{w}_j) \tag{11}$$

where $\mathbf{s}_t$ is chosen from an alphabet with 2^r elements for any $t \geqslant 0$, i.e., $\mathbf{s}_t$ can be represented by r bits. The data rate r of the symbols generated by the adaptive coder satisfies

$$\big|\log(\det(\boldsymbol{F}))\big| = \big|\log(\det(\boldsymbol{J}))\big| < r < \check{C}(\alpha)\,. \tag{12}$$

Symbol $\mathbf{s}_t$ is used by the anytime encoder as input at time t.

(d) Anytime encoder: Since $r < \check{C}(\alpha)$, there exists an anytime encoder-decoder pair that is α-reliable (see Section II). In particular, the anytime encoder generates message $\mathbf{m}_t$ at time t using $\mathbf{s}_{0:t}$ and this message is transmitted via the channel.

Receiver Blocks:

(e) Anytime decoder: Using received messages $\mathbf{r}_{0:t}$, the anytime decoder evaluates an estimator $\hat{\mathbf{s}}_{t'}(\mathbf{r}_{0:t})$ of $\mathbf{s}_{t'}$ for $t' \in \{0:t\}$. According to the definition of anytime reliability, (4) holds for a constant K that does not depend on t or t'.

(f) Internal estimator: At time t, an estimator $\hat{\mathbf{w}}_j(\mathbf{r}_{0:t})$ of $\mathbf{w}_j$ is evaluated as $\hat{\mathbf{w}}_j(\mathbf{r}_{0:t}) = \boldsymbol{f}_{\mathrm{s}}^{-1}\big(\hat{\mathbf{s}}_{j\tau:(j+1)\tau-1}(\mathbf{r}_{0:t})\big)$ for $j \in \{0:\lfloor (t+1)/\tau \rfloor - 1\}$, where $\boldsymbol{f}_{\mathrm{s}}^{-1}(\cdot)$ is the inverse of $\boldsymbol{f}_{\mathrm{s}}(\cdot)$ in (11). Combining the fact that $\boldsymbol{f}_{\mathrm{s}}^{-1}$ is a bijection with (4) gives

$$\mathbb{P}\big\{\hat{\mathbf{w}}_{0:j}(\mathbf{r}_{0:t}) \neq \mathbf{w}_{0:j}\big\} \leqslant K 2^{-\alpha(t-(j+1)\tau+1)} \tag{13}$$

where $\hat{\mathbf{w}}_{0:j}(\mathbf{r}_{0:t})$ represents the concatenation of $\hat{\mathbf{w}}_0(\mathbf{r}_{0:t})$, $\hat{\mathbf{w}}_1(\mathbf{r}_{0:t}), \ldots, \hat{\mathbf{w}}_j(\mathbf{r}_{0:t})$. Based on $\hat{\mathbf{w}}_{0:\lfloor (t+1)/\tau \rfloor - 1}(\mathbf{r}_{0:t})$, an estimator $\hat{\boldsymbol{\varphi}}_{t'}(\mathbf{r}_{0:t})$ of $\hat{\boldsymbol{\varphi}}_{\mathrm{q},t'}$ and an estimator $\hat{\mathsf{l}}_{t'}(\mathbf{r}_{0:t})$ of $\mathsf{l}_{t'}$ are computed according to (22) in Appendix A for $t' \in \{0:t-1\}$. Finally, an estimator $\hat{\boldsymbol{\psi}}_t$ of $\boldsymbol{\psi}_t$ is computed as

$$\hat{\boldsymbol{\psi}}_t := \sum_{t'=0}^{t-1} \boldsymbol{J}^{t-t'} \hat{\boldsymbol{\varphi}}_{t'}(\mathbf{r}_{0:t})\,. \tag{14}$$

(g) Linear transformation via $\boldsymbol{M}$: An estimator $\hat{\boldsymbol{\theta}}_t$ of $\boldsymbol{\theta}_t$ is obtained via a linear transformation $\hat{\boldsymbol{\theta}}_t := \boldsymbol{M}\hat{\boldsymbol{\psi}}_t$.

The signal processing performed by the transmitter and the receiver are summarized in Algorithm 1 and Algorithm 2, respectively. Appendix B presents a sketch of the proof for (2) when the above transmitter and receiver are employed. A detailed proof can be found in [28, Appendix B.2.1]. □

Remark 1: In Proposition 1, the magnitudes of eigenvalues of $\boldsymbol{F}$ are assumed to be no smaller than 1 for the simplicity of presentation. This proposition can be extended to general cases without this assumption by decomposing the state space into a stable and an unstable subspace via a linear transformation [15].

Proposition 1 is built on results in [16] and [17]. In particular, a control under communication constraints problem is considered in [16], where the aim is to stabilize a linear system in the mean square sense based on data generated by a coder with a limited rate. In [16], the coder performs adaptive vector quantization based on the system state, and an estimator of the state is evaluated using the quantization output. The control signal generated by the controller is a linear function of the state estimator and is thus negatively affected by the error of the estimator due to quantization. Such error is periodically corrected by the coder in order to stabilize the system.

Different from the control problem studied in [16], this letter considers filtering over noisy channels. Specifically, the estimation error is due not only to quantization or other types of lossy compression at the transmitter, but also to the corruption of received messages. In particular, error due to such corruption cannot be corrected at the transmitter as the channel output is only available to the receiver. As a result, an uncorrected error in decoding a data symbol generated by the transmitter at a past time can propagate to the estimator of the current state. Therefore, channel coding techniques are needed to protect the transmitted data symbols. This can be achieved by using anytime encoding and decoding. In particular, if the anytime capacity of the channel is larger than the threshold we found, then a pair of anytime encoder and decoder can be constructed such that the probability of error for decoding data symbols generated in the past is small. This compensates for the effects of error propagation and ensures that the estimation error is bounded.

We comment on the tightness of the sufficient condition presented in Proposition 1 by comparing it with a necessary condition for the estimation error to be bounded. Consider a mild assumption that the differential entropy of $\boldsymbol{\nu}_t$ is larger than some constant $\underline{h} > -\infty$ for all t. Then, there exist an encoder at the transmitter and an estimator $\hat{\boldsymbol{\theta}}_t$ of $\boldsymbol{\theta}_t$ at the receiver that achieve (2) only if the Shannon capacity C of the channel satisfies $C > \log(|\det(\boldsymbol{F})|)$. Note that the right-hand side of this inequality is the same as that in (6). Therefore, for a channel whose Shannon capacity equals its anytime capacity, such as a noiseless digital channel or a Gaussian channel with feedback [17], Proposition 1 presents a necessary and sufficient condition indeed. For a channel whose Shannon capacity and anytime capacity are different, such a difference determines the tightness of the condition in Proposition 1.

Next, Proposition 1 is extended to scenarios where the transmitter does not know the state perfectly and only has a noisy observation $\tilde{\boldsymbol{\theta}}_t$ of $\boldsymbol{\theta}_t$ at time t. Since the state is unknown, the transmitter generates the encoded message $\mathbf{m}_t$ based on $\tilde{\boldsymbol{\theta}}_{0:t}$. The next corollary shows that if the moment of $\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t\|$ is bounded over time, then (6) is still a sufficient condition for the existence of a transmitter and a receiver that achieve (2).

Corollary 1: Consider a scenario where $\mathbf{m}_t$ is generated based on $\tilde{\boldsymbol{\theta}}_{0:t}$, where $\tilde{\boldsymbol{\theta}}_t$ is a noisy observation of $\boldsymbol{\theta}_t$. If $\tilde{\boldsymbol{\theta}}_t$ satisfies

$$\sup_{t\geqslant 0} \mathbb{E}\Big\{\big\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t\big\|^a\Big\} < \infty \tag{15}$$

for some $a > b$, and (6) holds for the α-anytime capacity $\breve{C}(\alpha)$ of the channel with α satisfying (5), then there exist a transmitter and a receiver that achieve (2).

Proof: The idea of the proof is presented here due to limitation of space. Inequality (15) indicates that $\tilde{\boldsymbol{\theta}}_t$ can be written as $\tilde{\boldsymbol{\theta}}_t = \boldsymbol{F}\tilde{\boldsymbol{\theta}}_{t-1} + \tilde{\boldsymbol{\nu}}_t$ for some random vector $\tilde{\boldsymbol{\nu}}_t$ that satisfies (3) with $\boldsymbol{\nu}_t$ replaced by $\tilde{\boldsymbol{\nu}}_t$. Viewing $\tilde{\boldsymbol{\theta}}_t$ as the state, the transmitter and the receiver perform the processing described in the proof of Proposition 1 for generating the encoded messages and evaluating the estimator $\hat{\boldsymbol{\theta}}_t$, respectively. The estimator $\hat{\boldsymbol{\theta}}_t$ can be shown to satisfy (2). □

A remaining question is, in what scenarios can the transmitter obtain $\tilde{\boldsymbol{\theta}}_t$ satisfying (15)? An example scenario is that $\tilde{\boldsymbol{\theta}}_t$ is a local estimator of $\boldsymbol{\theta}_t$ at the transmitter based on its measurements of the state. In particular, consider the scenario where the transmitter obtains a measurement $\boldsymbol{\xi}_t$ at each time t given by $\boldsymbol{\xi}_t = \boldsymbol{G}\boldsymbol{\theta}_t + \boldsymbol{\eta}_t$, where $\boldsymbol{G}$ is a deterministic matrix and $\boldsymbol{\eta}_t$ is a zero-mean random vector representing the measurement noise. If $\boldsymbol{\eta}_t$ satisfies (3) with $\boldsymbol{\nu}_t$ replaced by $\boldsymbol{\eta}_t$ and if $(\boldsymbol{G}, \boldsymbol{F})$ is observable, then the transmitter can construct a local estimator $\tilde{\boldsymbol{\theta}}_t$ of $\boldsymbol{\theta}_t$ using $\boldsymbol{\xi}_{0:t}$ such that (15) holds. Note that $(\boldsymbol{G}, \boldsymbol{F})$ is observable if and only if the rank of the observability matrix $\big[\boldsymbol{G}^{\mathrm{T}} \quad \boldsymbol{F}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}} \ \cdots \ \big(\boldsymbol{F}^{\mathrm{T}}\big)^{n-1}\boldsymbol{G}^{\mathrm{T}}\big]^{\mathrm{T}}$ equals the dimension n of $\boldsymbol{\theta}_t$ [1].

Finally, an application of Proposition 1 to a numerical example in [2] is presented. In this example, the state is a two-dimensional vector evolving according to (1) with $\boldsymbol{F} = \left[\begin{smallmatrix}1.25 & 0\\ 1 & 1.1\end{smallmatrix}\right]$, and $\boldsymbol{\nu}_t$ is a Gaussian random vector whose covariance matrix is time-invariant. The eigenvalues of $\boldsymbol{F}$ are 1.25 and 1.1. Therefore, $\rho(\boldsymbol{F}) = 1.25$ and $\log(|\det(\boldsymbol{F})|) = 0.46$. Moreover, (3) holds for any $a > 2$. Applying Proposition 1 with $b = 2$, and letting a approach infinity, we conclude that if there exists an $\alpha > 2.5$ such that the α-anytime capacity of the channel satisfies $\breve{C}(\alpha) > 0.46$ bits per channel use, then there exist a transmitter and a receiver such that (2) holds with $b = 2$, i.e., the MSE of $\hat{\boldsymbol{\theta}}_t$ is bounded over time.

IV. Conclusion

This letter investigated the problem of network filtering using messages received via a noisy channel with general models for state evolution and communication channels. Specifically, we derived a sufficient condition for the channel quality that ensures desirable filtering accuracy without assuming that the state disturbance and channel noise are Gaussian. We showed that if the anytime capacity of the channel is above a threshold determined by the state evolution model, then the estimation error at the receiver is bounded over time. Our results provide insights for the design of communication and inference algorithms for real-time applications.

APPENDIX A
DETAILS OF THE ADAPTIVE CODER-CONTROLLER AND INTERNAL ESTIMATOR

The adaptive coder (see Fig. 1) employs a vector quantizer consisting of n scalar quantizers, where n is the dimension of $\boldsymbol{\theta}_t$. In particular, a scalar quantizer is parameterized by quantities μ, ϱ, and ν, which determine the base, the length of intervals, and the number of intervals, respectively, of the quantizer. The quantization period τ described in Section III as well as parameters μ and ϱ are determined as follows. Choose μ that satisfies $|\det(\boldsymbol{F})| < \mu < 2^{\check{C}(\alpha)}$. Moreover, choose appropriate values for τ, $\varrho > \mu^{b/(a-b)}$, and $\xi > 1$ so that the following hold

$$n \log \xi + \frac{n}{\tau} \log \mu < \check{C}(\alpha) - \log\big(|\det(\boldsymbol{F})|\big) \tag{16a}$$

$$\frac{b}{\tau} \log c_{\varrho} + b\big(1 + \log_{\mu} \varrho\big) \log(\xi \bar{\lambda}) < \alpha \tag{16b}$$

where $c_{\varrho} := \max\left\{1, \frac{\mu\varrho}{2\mu-2}\right\}$ and $\bar{\lambda} := \rho(\boldsymbol{F}) = \rho(\boldsymbol{J})$. In particular, we can always make (16) hold by choosing τ large enough, ϱ close to $\mu^{b/(a-b)}$ enough, and ξ close to 1 enough.

The scalar quantizer partitions the real line into μ^{ν} intervals, including two intervals $(-\infty, -\varrho^{\nu-1})$ and $(\varrho^{\nu-1}, \infty)$ with infinite lengths, as well as $\mu^{\nu} - 2$ intervals with finite lengths. Details of the partition are presented in [16]. Each interval is assigned a distinct integer index. In particular, $(-\infty, -\varrho^{\nu-1})$ and $(\varrho^{\nu-1}, \infty)$ are assigned indices 0 and $\mu^{\nu} - 1$, respectively, whereas the other intervals are assigned indices $1, 2, \ldots, \mu^{\nu} - 2$. For any $x \in \mathbb{R}$, the scalar quantizer first determines the index $\omega_{\nu}(x)$ of the interval which x belongs to. Then the scalar quantizer evaluates the reconstruction value $q_{\nu}\big(\omega_{\nu}(x)\big)$, where $q_{\nu}(\cdot) : \{0 : \mu^{\nu} - 1\} \mapsto \mathbb{R}$ is given by

$$q_{\nu}(\omega) := \begin{cases} \text{midpoint of interval } \omega & \text{if } \omega \in \{1 : \mu^{\nu} - 2\} \\ \varrho^{\nu-1} + \kappa_{\nu}\big(\mu^{\nu} - 1\big) & \text{if } \omega = \mu^{\nu} - 1 \\ -\varrho^{\nu-1} - \kappa_{\nu}\big(0\big) & \text{if } \omega = 0 \end{cases}$$

with function $\kappa_{\nu}(\cdot) : \{0 : \mu^{\nu} - 1\} \mapsto \mathbb{R}$ defined as

$$\kappa_{\nu}(\omega) := \begin{cases} \text{half length of interval } \omega & \text{if } \omega \in \{1 : \mu^{\nu} - 2\} \\ \frac{\mu}{2\mu-2}\big(\varrho^{\nu} - \varrho^{\nu-1}\big) & \text{if } \omega = 0 \text{ or } \mu^{\nu} - 1. \end{cases}$$

To present details for the vector quantizer, we introduce a partition for a general n-dimensional vector $\boldsymbol{x}$. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ represent the distinct eigenvalues of $\boldsymbol{F}$ that are either real or have positive imaginary parts. Partition $\boldsymbol{x}$ as

$$\boldsymbol{x} = \Big[\big(\boldsymbol{x}^{(1)}\big)^{\mathrm{T}} \quad \big(\boldsymbol{x}^{(2)}\big)^{\mathrm{T}} \quad \cdots \quad \big(\boldsymbol{x}^{(m)}\big)^{\mathrm{T}}\Big]^{\mathrm{T}} \tag{17}$$

where $\boldsymbol{x}^{(k)}$ consists of n_k entries for $k \in \{1 : m\}$. Here, n_k is the algebraic multiplicity of λ_k if $\lambda_k \in \mathbb{R}$, otherwise n_k is twice the algebraic multiplicity of λ_k. For any input $\boldsymbol{x} \in \mathbb{R}^n$, the vector quantizer computes an n-dimensional classification vector $\boldsymbol{\omega}(\boldsymbol{x})$ by applying scalar quantization $\omega_{\tau_k}(\cdot)$ on each entry of $\boldsymbol{x}^{(k)}$. In other words,

$$\boldsymbol{\omega}(\boldsymbol{x}) := \Big[\big(\boldsymbol{\omega}(\boldsymbol{x})^{(1)}\big)^{\mathrm{T}} \quad \big(\boldsymbol{\omega}(\boldsymbol{x})^{(2)}\big)^{\mathrm{T}} \quad \cdots \quad \big(\boldsymbol{\omega}(\boldsymbol{x})^{(m)}\big)^{\mathrm{T}}\Big]^{\mathrm{T}} \tag{18}$$

with

$$\big[\boldsymbol{\omega}(\boldsymbol{x})^{(k)}\big]_i = \omega_{\tau_k}\big(\big[\boldsymbol{x}^{(k)}\big]_i\big), \qquad i \in \{1 : n_k\}. \tag{19}$$

Here, τ_k determines the number of intervals for the scalar quantizion applied on $\big[\boldsymbol{x}^{(k)}\big]_i$ and is given by

$$\tau_k := \big\lfloor \tau \log_{\mu}(\xi|\lambda_k|)\big\rfloor + 1\,.$$

The vector quantizer also evaluates a reconstruction vector $\boldsymbol{q}\big(\boldsymbol{\omega}(\boldsymbol{x})\big) \in \mathbb{R}^n$. In particular, the argument of function $\boldsymbol{q}(\cdot)$ is an n-dimensional vector $\boldsymbol{w}$ such that each entry of $\boldsymbol{w}^{(k)}$ belongs to $\{0 : \mu^{\tau_k} - 1\}$, and $\boldsymbol{q}(\boldsymbol{w})$ is obtained by applying $q_{\tau_k}(\cdot)$ on each entry of $\boldsymbol{w}^{(k)}$. In other words, the ith entry of sub-vector $\boldsymbol{q}(\boldsymbol{w})^{(k)}$ is given by $\big[\boldsymbol{q}(\boldsymbol{w})^{(k)}\big]_i = q_{\tau_k}\big(\big[\boldsymbol{w}^{(k)}\big]_i\big)$.

The adaptive coder evaluates an estimator $\hat{\boldsymbol{\varphi}}_{\mathrm{q},t}$ of $\boldsymbol{\varphi}_t$ and a scaling factor l_t at each time t as follows

$$\hat{\boldsymbol{\varphi}}_{\mathrm{q},t} = \begin{cases} \boldsymbol{J}^{\tau} \mathsf{l}_{j\tau}\, \boldsymbol{q}(\mathbf{w}_j) & \text{if } \frac{t}{\tau} = j+1 \text{ for } j \in \mathbb{N} \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{20a}$$

$$\mathsf{l}_t = \begin{cases} \max\Big\{\sigma, \max\limits_{\substack{k\in\{1:m\}\\ i\in\{1:n_k\}}} \big\{|\lambda_k|^{\tau} \kappa_{\tau_k}\big(\big[\mathbf{w}_{\ell}^{(k)}\big]_i\big)\big\}\mathsf{l}_{\ell\tau}\Big\} & \\ & \text{if } \frac{t}{\tau} = \ell+1 \text{ for } \ell \in \mathbb{N} \\ \mathsf{l}_{t-1} & \text{if } \frac{t}{\tau} \notin \mathbb{N} \\ \sigma & \text{if } t = 0 \end{cases} \tag{20b}$$

where $\mathbf{w}_{\ell}^{(k)}$ is a sub-vector of the classification vector $\mathbf{w}_{\ell}$ and has n_k entries (see (17)), and σ is a constant such that

$$\sup_{j\in\mathbb{N}} \mathbb{E}\Big\{\Big\|\sum_{l=1}^{\tau} \boldsymbol{J}^{\tau-l}\boldsymbol{M}^{-1}\mathbf{v}_{j\tau+l}\Big\|^a\Big\} \leqslant \sigma^a.$$

At time $j\tau$, the adaptive coder computes the classification vector $\mathbf{w}_j$ according to (10), where $\boldsymbol{\omega}(\cdot)$ is defined in (18) and (19). Vector $\mathbf{w}_j$ can be represented by $\sum_{k=1}^{m} n_k\tau_k \log \mu$ bits as $\mathbf{w}_j^{(k)}$ consists of n_k entries with each entry $\big[\mathbf{w}_j^{(k)}\big]_i \in \{0 : \mu^{\tau_k} - 1\}$. Since a classification vector is generated every τ time steps, the data rate r of the coder is $r = \frac{1}{\tau}\sum_{k=1}^{m} n_k\tau_k \log \mu$, which satisfies (12). The control $\mathbf{u}_t$ at time t is given by

$$\mathbf{u}_t := -\boldsymbol{J}\hat{\boldsymbol{\varphi}}_{\mathrm{q},t}\,. \tag{21}$$

Note that $\mathbf{u}_t$ is completely determined by quantization output as $\hat{\boldsymbol{\varphi}}_{\mathrm{q},t}$ is a function of classification vectors. If the above coder-controller is employed, then (9) holds.

Next, expressions of $\hat{\boldsymbol{\varphi}}_{t'}(\mathbf{r}_{0:t})$ and $\hat{\mathsf{l}}_{t'}(\mathbf{r}_{0:t})$ evaluated by the internal estimator are presented:

$$\hat{\boldsymbol{\varphi}}_{t'}(\mathbf{r}_{0:t}) = \begin{cases} \boldsymbol{J}^{\tau}\hat{\mathsf{l}}_{j\tau}(\mathbf{r}_{0:t})\, \boldsymbol{q}(\hat{\mathbf{w}}_j(\mathbf{r}_{0:t})) & \text{if } \frac{t'}{\tau} = j+1 \text{ for } j \in \mathbb{N} \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{22a}$$

$$\hat{\mathsf{l}}_{t'}(\mathbf{r}_{0:t}) = \begin{cases} \max\Big\{\sigma, \max\limits_{\substack{k\in\{1:m\}\\ i\in\{1:n_k\}}} \big\{|\lambda_k|^{\tau} \kappa_{\tau_k}\big(\big[\hat{\mathbf{w}}_{\ell}(\mathbf{r}_{0:t})^{(k)}\big]_i\big)\big\}\hat{\mathsf{l}}_{\ell\tau}(\mathbf{r}_{0:t})\Big\} & \\ & \text{if } \frac{t'}{\tau} = \ell+1 \text{ for } \ell \in \mathbb{N} \\ \hat{\mathsf{l}}_{t'-1}(\mathbf{r}_{0:t}) & \text{if } \frac{t'}{\tau} \notin \mathbb{N} \\ \sigma & \text{if } t' = 0 \end{cases} \tag{22b}$$

where $\hat{\mathbf{w}}_{\ell}(\mathbf{r}_{0:t})^{(k)}$ is a sub-vector of $\hat{\mathbf{w}}_{\ell}(\mathbf{r}_{0:t})$ and has n_k entries (see (17)). Comparing (20) with (22) gives

$$\hat{\boldsymbol{\varphi}}_{t'}(\mathbf{r}_{0:t}) = \hat{\boldsymbol{\varphi}}_{\mathrm{q},t'} \qquad \forall t' \in \{0 : (j+1)\tau - 1\}, \quad \text{if } \hat{\mathbf{w}}_{0:j-1}(\mathbf{r}_{0:t}) = \mathbf{w}_{0:j-1}\,. \tag{23}$$

APPENDIX B
SKETCH OF PROOF FOR (2)

Proof: Consider the following estimator $\hat{\boldsymbol{\psi}}_{\mathrm{q},t}$ of $\boldsymbol{\psi}_t$

$$\hat{\boldsymbol{\psi}}_{\mathrm{q},t} := \sum_{t'=0}^{t-1} \boldsymbol{J}^{t-t'} \hat{\boldsymbol{\varphi}}_{\mathrm{q},t'} \,. \tag{24}$$

Note that $\hat{\boldsymbol{\psi}}_{\mathrm{q},t}$ is a function of the quantization output of the adaptive coder. Combining (24) with (21) and (8) gives $\boldsymbol{\varphi}_t = \boldsymbol{\psi}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t}$, which shows $\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t = \boldsymbol{M}(\hat{\boldsymbol{\psi}}_t - \boldsymbol{\psi}_t) = \boldsymbol{M}(\hat{\boldsymbol{\psi}}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t} - \boldsymbol{\varphi}_t)$. Applying triangle inequality as well as the inequality $(\sum_{l=1}^{L} y_l)^x \leqslant L^x \sum_{l=1}^{L} y_l^x$ for any positive integer L and positive scalars $x, y_1, y_2, \ldots, y_L$, we obtain

$$\mathbb{E}\{\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t\|^b\} \leqslant \|\boldsymbol{M}\|^b 2^b \Big(\mathbb{E}\{\|\hat{\boldsymbol{\psi}}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t}\|^b\} + \mathbb{E}\{\|\boldsymbol{\varphi}_t\|^b\}\Big) \,. \tag{25}$$

Combining (25) with (9), in order to establish (2) we only need to show

$$\sup_{t\geqslant 0} \mathbb{E}\Big\{\|\hat{\boldsymbol{\psi}}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t}\|^b\Big\} < \infty \,. \tag{26}$$

The idea for proving (26) is described here and a detailed derivation is presented in [28, Appendix B.2.1]. According to (14) and (24), $\hat{\boldsymbol{\psi}}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t}$ is determined by the differences between $\hat{\boldsymbol{\varphi}}_{t'}(\mathbf{r}_{0:t})$ and $\hat{\boldsymbol{\varphi}}_{\mathrm{q},t'}$ for $t' \in \{0 : t-1\}$. According to (22) and (20), such differences are due to errors in $\hat{\mathbf{w}}_j(\mathbf{r}_{0:t})$ for $j \in \{0 : j_t - 1\}$, where $j_t := \lfloor (t-1)/\tau \rfloor$. Define $\mathrm{j_e}$ as the smallest integer $j \in \{0 : j_t - 1\}$ such that $\hat{\mathbf{w}}_j(\mathbf{r}_{0:t}) \neq \mathbf{w}_j$. If such j does not exist, then set $\mathrm{j_e} = j_t$. By definition, $\hat{\mathbf{w}}_{0:\mathrm{j_e}-1}(\mathbf{r}_{0:t}) = \mathbf{w}_{0:\mathrm{j_e}-1}$. According to (23),

$$\hat{\boldsymbol{\varphi}}_{t'}(\mathbf{r}_{0:t}) = \hat{\boldsymbol{\varphi}}_{\mathrm{q},t'}, \qquad \forall t' \in \{0 : (\mathrm{j_e}+1)\tau - 1\} \,. \tag{27}$$

Combining (14), (24), (22a), (20a), and (27) gives

$$\hat{\boldsymbol{\psi}}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t} = \sum_{j=\mathrm{j_e}+1}^{j_t} \boldsymbol{J}^{t-j\tau} \big(\hat{\boldsymbol{\varphi}}_{j\tau}(\mathbf{r}_{0:t}) - \hat{\boldsymbol{\varphi}}_{\mathrm{q},j\tau}\big) \,. \tag{28}$$

Equation (28) shows that the difference $\hat{\boldsymbol{\varphi}}_{(\mathrm{j_e}+1)\tau}(\mathbf{r}_{0:t}) - \hat{\boldsymbol{\varphi}}_{\mathrm{q},(\mathrm{j_e}+1)\tau}$, caused by error in estimating $\mathbf{w}_{\mathrm{j_e}}$, is multiplied by $\boldsymbol{J}^{t-(\mathrm{j_e}+1)\tau}$ when we compute $\hat{\boldsymbol{\psi}}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t}$. Therefore, the effect of the error in estimating $\mathbf{w}_{\mathrm{j_e}}$ on $\hat{\boldsymbol{\psi}}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t}$ would increase exponentially with $t - (\mathrm{j_e}+1)\tau$ if the eigenvalues of $\boldsymbol{J}$ have magnitudes larger than 1. On the other hand, since anytime encoder and decoder are employed, the probability that there is an error in estimating $\mathbf{w}_{\mathrm{j_e}}$ decreases exponentially with $t - (\mathrm{j_e}+1)\tau$ (see (13)). When (5) and (16b) are satisfied, the exponential decrease of the probability of error in estimating $\mathbf{w}_{\mathrm{j_e}}$ outweighs the exponential increase of the error's effect on $\hat{\boldsymbol{\psi}}_t - \hat{\boldsymbol{\psi}}_{\mathrm{q},t}$. As a result, (26) holds. Combining (9), (25), and (26) gives (2). □

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[2] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[3] T. Tanaka, M. Skoglund, and V. Ugrinovskii, "Optimal sensor design and zero-delay source coding for continuous-time vector Gauss-Markov processes," in *Proc. IEEE Conf. Decis. Control*, Melbourne, VIC, Australia, Dec. 2017, pp. 3992–3997.

[4] M. Z. Win *et al.*, "Network localization and navigation via cooperation," *IEEE Commun. Mag.*, vol. 49, no. 5, pp. 56–62, May 2011.

[5] A. Conti *et al.*, "Location awareness in beyond 5G networks," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 22–27, Nov. 2021.

[6] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola, and C. Fischione, "A survey of enabling technologies for network localization, tracking, and navigation," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3607–3644, 4th Quart., 2018.

[7] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient multisensor localization for the Internet of Things: Exploring a new class of scalable localization algorithms," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 153–167, Sep. 2018.

[8] A. A. Saucan and M. Z. Win, "Information-seeking sensor selection for ocean-of-things," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10072–10088, Oct. 2020.

[9] M. Chiani, A. Giorgetti, and E. Paolini, "Sensor radar for object tracking," *Proc. IEEE*, vol. 106, no. 6, pp. 1022–1041, Jun. 2018.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[11] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 172–182, Apr. 1966.

[12] N. J. Newton and S. K. Mitter, "Variational Bayes and a problem of reliable communication I: Finite systems," *Commun. Inf. Syst.*, vol. 10, no. 3, pp. 155–182, Aug. 2010.

[13] N. J. Newton and S. K. Mitter, "Variational Bayes and a problem of reliable communication: II. Infinite systems," *J. Stat. Mech. Theory Exp.*, vol. 2012, no. 11, pp. 1–30, Nov. 2012.

[14] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 6–17, Jan. 2010.

[15] S. Tatikonda and S. K. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1056–1068, Jul. 2004.

[16] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM J. Control Optim.*, vol. 43, no. 2, pp. 413–436, 2004.

[17] A. Sahai and S. K. Mitter, "The necessity and sufficiency of anytime capacity for Stabilization of a linear system over a noisy communication link—Part I: Scalar systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3369–3395, Aug. 2006.

[18] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information feedback," *IEEE Trans. Autom. Control*, vol. 50, no. 6, pp. 910–915, Jun. 2005.

[19] P. Minero, M. Franceschetti, S. Dey, and G. N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 243–255, Feb. 2009.

[20] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*. New York, NY, USA: Springer, 2013.

[21] A. Conti, S. Mazuelas, S. Bartoletti, W. C. Lindsey, and M. Z. Win, "Soft information for localization-of-things," *Proc. IEEE*, vol. 107, no. 11, pp. 2240–2264, Nov. 2019.

[22] S. Kar and J. M. F. Moura, "Consensus + innovations distributed inference over networks: Cooperation and sensing in networked systems," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 99–109, May 2013.

[23] S. Marano, V. Matta, and P. Willett, "Asymptotic design of quantizers for decentralized MMSE estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5485–5496, Nov. 2007.

[24] Z. Liu, A. Conti, S. K. Mitter, and M. Z. Win, "Networked filtering with feedback for discrete-time observations," in *Proc. IEEE Conf. Decis. Control*, Austin, TX, USA, Dec. 2021, pp. 5882–5889.

[25] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*. New York, NY, USA: Cambridge Univ. Press, 2018.

[26] P. Minero and M. Franceschetti, "Anytime capacity of a class of Markov channels," *IEEE Trans. Autom. Control*, vol. 62, no. 3, pp. 1356–1367, Mar. 2017.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2013.

[28] Z. Liu, "Decentralized inference and its application to network localization and navigation," Ph.D. dissertation, Dept. Aeronaut. Astronaut., Massachusetts Inst. Technol., Cambridge, MA, USA, May 2022.