Towards a Data-Rate Theorem for State Estimation with Guaranteed Accuracy in Bounded-Noise Environments

1st Guilherme S. Vicinansa
Department of Electrical and Electronic Engineering,
The University of Melbourne
Melbourne, Australia
guilherme.vicinansa@unimelb.edu.au

2nd Girish N. Nair
Department of Electrical and Electronic Engineering,
The University of Melbourne
Melbourne, Australia
gnair@unimelb.edu.au

***Abstract*—In this work, we analyze the minimum data-rate to reconstruct the state trajectory of a perturbed discrete-time linear system with a prescribed accuracy. The data is transmitted periodically at every sample. Among the results, we prove data-rate lower and upper bounds that grow to infinity as a "disturbance-to-accuracy ratio" approaches one. In general, when considered per dimension, the gap between the upper and lower bounds grows at most logarithmically and, in many cases, remains constant. Furthermore, we discuss why such a result must be coordinate-dependent, a feature that differs from the noiseless case. Finally, we illustrate our results with an example.**

***Index Terms*—Network control, intrinsic entropy, state estimation**

I. Introduction

Control systems with distributed components, such as sensors and actuators, are prevalent in practical applications. Those components often need to share measurement results, which requires communication channels. These channels, in turn, limit the amount of data the transmitter can send over a given time interval to a finite number of symbols. This latter restriction, on the other hand, limits the accuracy with which the receiver can reconstruct the original message. Thus, understanding the relationship between this accuracy and the communication data-rate is relevant to solving distributed control and estimation problems efficiently.

It is now well-established that there is a minimum transmission data-rate below which some control [1], [5], [9], [16], [17] and estimation ([6], [13], [15], [22], [27]) problems have no solution. In particular, in the absence of perturbations, we know that the minimum data-rate to reconstruct the state with a prescribed uniform upper bound for the estimation error is the topological entropy of the system [22]. When perturbations are considered, the work [21] provides a lower bound for such a minimal data-rate for nonlinear systems that satisfy a type of dissipation inequality. However, data-rate theorems providing upper and lower bounds for state estimation of perturbed linear systems with accuracy guarantees are missing from the literature. The goal of the present document is to fill that gap.

In general, there is a maximal time for the transmitter to send the data to keep the state estimation error within a prescribed bound. The reason is that if we wait too long to transmit, the cumulative effect of the disturbance eventually increases the set of possible states to a point where it is too large, making it impossible to build an estimator that is guaranteed to satisfy the desired accuracy bound. Consequently, we cannot use asymptotic techniques, such as those presented in [13], [22], [27], [29], to analyze this problem. The present results show that the data-rate in noisy environments must be higher than that usually provided by classical data-rate theorems. Remarkably, we prove that when the effect of the disturbance becomes comparable to the required accuracy, the minimum data-rate to solve this problem grows unbounded. We arrive at this conclusion by providing a data-rate lower bound below which the state reconstruction problem has no solution. Additionally, we show that the problem always has a solution if a"disturbance-to-accuracy ratio" is smaller than one.

It is worth mentioning that the work [21] also studies the problem of reconstructing the state of a nonlinear system under accuracy requirements and the effect of noise. Their result also utilizes a disturbance-to-accuracy ratio in its lower bound. Nevertheless, our work differs from that in a few aspects. First, although their result is stated for nonlinear systems, [21] does not provide a data-rate upper bound. In the present work, we restrict our attention to linear systems, a fact that allows us to provide both lower and upper bounds. Moreover, we show that the data-rate gap between the two bounds grows at most logarithmically and, when the accuracy is similar in size to the noise, is constant when considered per dimension, i.e., when dividing the gap by the state dimension. Second, as mentioned above, our lower bound grows unbounded when the disturbance-to-accuracy is close to one, a new feature not explicitly captured by the result in [21]. Third, our proof methods, presented in the report [**?**], are novel and might provide new tools, complementary to those in [21], for deriving new data-rate theorems.

This document is structured as follows: in Section II, we informally pose and motivate the problem. Also, we present a motivating example that illustrates the types of results obtained in this paper. Then, in Section III, we introduce the uncertain model for the class of systems we study. This model provides a way to understand how the perturbed system dynamics generates uncertainty. We do that using the notion of uncertain variables introduced in [17]. We also formally state the problem we want to study. Then, in Section IV, we present our main results: we claim that the state estimation problem has a solution, and we provide data-rate lower and upper bounds. Next, we analyze and illustrate the results by revisiting the motivating example. Finally, we present our conclusions.

Notations: let $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{Z}_{\geq 0}$, and $\mathbb{Z}_{>0}$ denote the sets of real, integer, nonnegative integer, and positive integer numbers, respectively. Given integers $a < b$, and $c > 0$, denote by $[a:b] := \{a, \dots, b\}$ and by $[c] := \{1, \dots, c\}$. Given a set $\mathcal{A}$, we define $a^{0:n} := (a_0, \dots, a_n) \in \mathcal{A}^{n+1}$. For a vector $x \in \mathbb{R}^d$, we denote by $|x|$ its Euclidean norm. A function $f : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ is Big-O of $g : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ if $\limsup_{n\to\infty} f(n)/g(n) \leq \infty$. Given $r \in \mathbb{R}$, we denote by $\lfloor r \rfloor$ and $\lceil r \rceil$ its floor and ceiling values, respectively. We denote by ω_k the k-dimensional Euclidean unit ball's volume. Finally, we denote by $\log(x)$ the logarithm in base 2 of $x \in (0, \infty)$.

II. The problem and motivation

In this work, we are interested in studying fundamental limitations on the data-rate associated with the problem of state reconstruction for discrete-time linear systems subject to the effect of unknown disturbances. Specifically, our goal is to understand what is the minimum data-rate associated with keeping the estimation error small when disturbances are present. We begin this session by introducing the model and by giving an informal description of the problem we want to address. Then, we present a motivating example that helps illustrate the theory.

Consider a perturbed discrete-time linear time-invariant (LTI) system. Explicitly, for each $n \in \mathbb{Z}_{\geq 0}$, we have that

$$x_{n+1} = Ax_n + Bu_n + v_n, \tag{1}$$

where $x_0 \in \mathbb{R}^d$ is the initial state, $A \in \mathbb{R}^{d\times d}$ and $B \in \mathbb{R}^{d\times m}$ are matrices, $(v_n)_{n\in\mathbb{Z}_{\geq 0}}$ is a sequence of unknown disturbances, and $(u_n)_{n\in\mathbb{Z}_{\geq 0}}$ is a sequence of known control actions. Also assume that A is invertible.

The informal description of our problem is as follows: for each time $n \in \mathbb{Z}_{\geq 0}$, we want to build an estimate $\hat{x}_n$ for the value of x_n. However, the estimator and the plant are physically far from each other. This forces us to transmit the state measurements over a communication channel, which we assume is a digital channel. Since digital channels can only transmit a finite number of symbols per unit of time, we must encode the measurement using a finite alphabet $\mathcal{A}$. This implies that the channel transmits data at an rate of $\log(\#\mathcal{A})$ bits per sample. Since the set of possible packets the estimator can receive from the channel at each time n is finite, we note that, even if we measure the state perfectly, the estimation error cannot be zero in general. Clearly, if the estimator knew x_0 exactly beforehand and there were no disturbances, it could use the recursion $\hat{x}_{n+1} = A\hat{x}_n + Bu_n$ with $\hat{x}_0 = x_0 \in \mathbb{R}^d$ to reconstruct the state since the estimator knows the control u_n. Unfortunately, the estimator does not know x_0 and there are disturbances affecting our system dynamics. We assume, nevertheless, that the estimator knows an estimate $p_0 \in \mathbb{R}^d$ and constants $\gamma \in [0, \infty)$ and $\varepsilon_0 \in (0, \infty)$ so that $|v_n| \leq \gamma$ and that $|x_0 - p_0| \leq \varepsilon_0$, i.e, it knows a bound for the noise and an approximation for the initial state. Note that, we can take $\hat{x}_0 = p_0$ and guarantee that $|x_0 - \hat{x}_0| \leq \varepsilon_0$. Thus, as an alternative goal to perfect state reconstruction, we require the estimator to keep the estimation error uniformly bounded by some prescribed accuracy level $\varepsilon \geq \varepsilon_0$ for all times. We know from the literature [22] that this can be done if $v_n = 0$ for all $n \in \mathbb{Z}_{\geq 0}$ when ε_0 is small enough.

This raises some questions: does there exist a minimum data-rate above which such an estimator exists? Assuming this problem has a solution, can we provide a simple lower bound for the minimum data-rate below which this problem has no solution and explain how it varies with ε_0, ε, and γ? We answer both questions affirmatively in Theorem 1. In that theorem, we provide upper and lower bounds for the minimum data-rate. Also, we remark that the expressions obtained depend only on the noise-to-accuracy ratio γ/ε and on the dynamics, i.e., on A.

To motivate our results, consider the following example. We revisit this example in Section IV once we developed the theory.

Example 1 (Matrices with the same intrinsic entropy) Assume that the initial accuracy and the desired accuracy are the same $\varepsilon_0 = \varepsilon$. Our goal is to understand how the minimum data-rate for estimating the state of system (1) with a uniform estimation error bound varies as a function of the disturbance level γ, the accuracy ε, and the the system matrix A. To do that, consider matrices $A_1 = \begin{pmatrix} 0.9 & 2 \\ 0 & 0.5 \end{pmatrix}$ and $A_2 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.5 \end{pmatrix}$. Both matrices are Schur stable, i.e., their eigenvalues lie inside the unit circle in the complex plane. We prove in Theorem 1 that the minimum data-rate for reconstructing the state of system (1) with an ε accuracy for each matrices above is lower-bounded by functions that are $O(\log(1/(1-r)))$ with $r = \gamma/\varepsilon$. Intuitively, this fact shows us that if the disturbance is similar in magnitude to the accuracy, the minimum data-rate becomes large. Also, Theorem 1 gives different minimum data-rates for A_1 and A_2. We analyze this observation in Subsection V-B. ▲

Model (1) describes the state evolution for a given initial state x_0, a given sequence of disturbances $(v_n)_{n\in\mathbb{Z}_{\geq 0}}$, and a given sequence of controls $(u_n)_{n\in\mathbb{Z}_{\geq 0}}$. To formally state and study our problem, we need a model that relates how the uncertainty in the sequence of disturbances and in the initial condition affects the uncertainty in the state. This is the goal of the next section.

III. MODELING UNCERTAINTY

Consider the following informal discussion: at time n, the estimator knows that the state belongs to a known set, a fact it inferred from past received data. For this reason, we call this set the set of possible states at time n with prior knowledge. Without new information, the estimator's uncertainty about the state would grow since the system dynamics would make the set of possible states at time $n+1$ larger. The encoder, in turn, transmits data that carries information about the state at time n. In this manner, the estimator can update its uncertainty about the state at time n and obtain a set of possible states with posterior knowledge that is "more accurate" than the set with prior knowledge. Finally, the estimator can infer where the state would be at time $n+1$ using the set of possible states at time n and the system dynamics. Thus, in some sense, we can think that the system dynamics generate uncertainty while the symbols the encoder transmits reduce it. Therefore, to find the minimal data-rate for solving our problem, we must understand how fast system (1) generates uncertainty.

The goal of this section is to introduce tools that help us formalize the above discussion. First, we recall the notion of uncertain variables [17] and show how we can use it to model how the uncertainty in system (1) evolves. Then, we formally state the informal assumptions and goals described in the previous section using the uncertain variable framework. Finally, we pose the estimation problem in this setting.

A. Uncertain variables

Our first step to formally describe how system (1) propagates uncertainty over time is to recall the notion of uncertain variables.

Definition 1 (Uncertain variables [17])
Let $(\Omega, \mathscr{F})$ be a measurable space. We call Ω the *sample space* and any element $\omega \in \Omega$ is called a *sample*. A Borel measurable map $\mathbf{X} : \Omega \to \mathcal{X}$ is called an *uncertain variable* (uv). We define an uncertain variable *(marginal) range* by $[\![\mathbf{X}]\!] \coloneqq \{\mathbf{X}(\omega) \in \mathcal{X} : \omega \in \Omega\}$. Also, given another uncertain variable $\mathbf{Y} : \Omega \to \mathcal{Y}$, we define the *conditional range of* $\mathbf{Y}$ *given* $\mathbf{X} = x$ as $[\![\mathbf{Y}|\mathbf{X} = x]\!] \coloneqq \{\mathbf{Y}(\omega) \in \mathcal{Y} : \omega \in \Omega \text{ and } \mathbf{X}(\omega) = x\}$. Further, we say that $\mathbf{Y}$ and $\mathbf{X}$ are *unrelated*, which we denote by $\mathbf{Y} \perp\!\!\!\perp \mathbf{X}$, if $[\![\mathbf{Y}|\mathbf{X} = x]\!] = [\![\mathbf{Y}]\!]$ for all $x \in [\![\mathbf{X}]\!]$. ▲

To better understand how uvs can help our discussion, consider the following example: the estimator can model the possible initial states as a uv $\mathbf{X}_0$. In this case, the true initial state is represented by $\mathbf{X}_0(\omega)$ for some unknown $\omega \in \Omega$, and the set $[\![\mathbf{X}_0]\!]$ corresponds to the set of possible initial states. Now, we can model the encoder as another uv $\mathbf{\Gamma}_0$. Similarly to the case of the initial state, we should have that $\mathbf{\Gamma}_0(\omega) = a_0$ corresponds to the encoded value of $\mathbf{X}_0(\omega)$ and that the set of possible symbols is contained in the coder alphabet, i.e., $[\![\mathbf{\Gamma}_0]\!] \subseteq \mathcal{A}$. Thus, even though the estimator does not know ω, it knows $\mathbf{\Gamma}_0(\omega) = a_0$, the encoded value of $\mathbf{X}_0(\omega)$. Hence, we need a way to model how receiving the information that $\mathbf{\Gamma}_0(\omega) = a_0$ affects the estimator uncertainty on $\mathbf{X}_0$. To do that, we use the conditional range $[\![\mathbf{X}_0|\mathbf{\Gamma}_0 = a_0]\!]$, which is the subset of elements of $[\![\mathbf{X}_0]\!]$ that are encoded as a_0, i.e., $[\![\mathbf{X}_0|\mathbf{\Gamma}_0 = a_0]\!]$ is our a posteriori knowledge about $\mathbf{X}_0(\omega)$ given a_0. Therefore, this framework gives us a concise and clear way of describing how new information affects the estimator uncertainty. Nonetheless, we still need to comprehend how the dynamics (1) generates uncertainty. To understand that is the goal of our next subsection.

B. The problem setting and the uncertain model

Denote by $\mathbf{X}_n$, $\hat{\mathbf{X}}_n$, $\mathbf{V}_n$, $\mathbf{U}_n$, and $\mathbf{\Gamma}_n$ the uvs representing the state, the state estimate, the disturbance, the control, and the encoder at time $n \in \mathbb{Z}_{\geq 0}$, respectively. With this notation, we can convert the informal setting described in Section II into properties of the uvs. First, we explain how our coding, decoding, and controller work: there exist sequences of functions $(\gamma_n)_{n \in \mathbb{Z}_{\geq 0}}$, $(c_n)_{n \in \mathbb{Z}_{\geq 0}}$, and $(\delta_n)_{n \in \mathbb{Z}_{\geq 0}}$ so that

$$\mathbf{\Gamma}_0 = \gamma_0(\mathbf{X}_0) \text{ and } \mathbf{\Gamma}_n = \gamma_n(\mathbf{X}_n, \mathbf{\Gamma}_0, \ldots, \mathbf{\Gamma}_{n-1}) \tag{2a}$$

$$\mathbf{U}_n = c_n(\mathbf{\Gamma}_0, \ldots, \mathbf{\Gamma}_{n-1}) \tag{2b}$$

$$\hat{\mathbf{X}}_n = \delta_n(\mathbf{\Gamma}_0, \ldots, \mathbf{\Gamma}_{n-1}) \tag{2c}$$

for each $n \in \mathbb{Z}_{>0}$. Here, we convention that $\mathbf{U}_0$ and $\hat{\mathbf{X}}_0$ are decided before operation. Equation (2a) means that the symbol used to encode the state at time n is a function of the past symbols and the current state, i.e., it has memory. These coders are well-studied and used in practice (see, e.g., predictive quantizers in Chapter 7 from [8]). Equations (2b) and (2c) imply that the control action and the state estimate are functions of the previously received symbols, respectively. We also impose some mild additional technical constraints.

Assumption 1

i) $[\![\mathbf{X}_0]\!] = \mathbb{B}(p_0; \varepsilon)$ where p_0 is known;

ii) $[\![\mathbf{V}_n]\!] = \mathbb{B}(0; \gamma)$, $\mathbf{V}_n \perp\!\!\!\perp \mathbf{X}_k$, and $\mathbf{V}_n \perp\!\!\!\perp \mathbf{\Gamma}_k$ for all $k \in [0 : n]$ and all $n \in \mathbb{Z}_{\geq 0}$;

iii) $[\![\mathbf{X}_n|\mathbf{\Gamma}_0 = a_0, \ldots, \mathbf{\Gamma}_n = a_n]\!]$ is a Borel set with nonempty interior for each $n \in \mathbb{Z}_{\geq 0}$. ▲

Some observations are in order. First, i) implies that we know an initial state estimate with accuracy ε. Second, condition iii) is less clear and requires some explaining: $[\![\mathbf{X}_n|\mathbf{\Gamma}_0 = a_0, \ldots, \mathbf{\Gamma}_n = a_n]\!]$ is the set of possible states at time n after receiving the symbols $(a_0, \ldots, a_n)$. We interpret this set as a quantization region for a quantizer with memory, i.e., the encoder quantizes the state. In the proof of the data-rate lower bound, we perform a k-dimensional volume counting, where $k \in [d]$. For that reason, we need to impose this measurability condition. Recall that all open sets, closed sets, and their countable unions and intersections are Borel sets (see, e.g., Section 1.2 from [7]). For example, half-open parallelepipeds, i.e., sets $\{\sum_{i=1}^{d} \beta_i e_i : \beta_i \in [\underline{a}_i, \overline{a}_i)\}$ with $\underline{a}_i < \overline{a}_i$ for each $i \in [d]$, which are common shapes for quantization regions, are Borel sets. We also remark that building sets that are not Borel requires some effort (see, e.g., Section 1.5 from [7]).

Next, we want to rewrite equation (1) using uvs. We do that as follows.

$$\mathbf{X}_{n+1} = A\mathbf{X}_n + \mathbf{U}_n + \mathbf{V}_n \tag{3}$$

and refer to is as the *uncertain model* for the dynamics (1).

We briefly remark that the uncertain variable approach is not the only possible way of modeling the kind of uncertainty we consider here. We refer the reader to [3], [4], [11], [23], [24], [28] for alternative choices and to [**?**] for a thorough discussion on why we chose this setting.

C. The coder-estimator scheme and the problem statement

Before we mathematically formalize our problem statement, we make a definition.

Definition 2
A *state estimator* is a sequence of uvs $(\hat{\mathbf{X}}_n)_{n\in\mathbb{Z}_{\geq 0}}$ where the elements have range in $\mathbb{R}^d$. Given a state estimator $(\hat{\mathbf{X}}_n)_{n\in\mathbb{Z}_{\geq 0}}$ and a coder $(\mathbf{\Gamma}_n)_{n\in\mathbb{Z}_{\geq 0}}$ with alphabet $\mathcal{A}$, we say that $\mathscr{S} = \big((\hat{\mathbf{X}}_n)_{n\in\mathbb{Z}_{\geq 0}}, (\mathbf{\Gamma}_n)_{n\in\mathbb{Z}_{\geq 0}}\big)$ is a *coder-estimator scheme*. We say that a coder-estimator scheme $\mathscr{S}$ is *ε-accurate* for $(\mathbf{X}_n)_{n\in\mathbb{Z}_{\geq 0}}$ if $[\![\mathbf{X}_n - \hat{\mathbf{X}}_n]\!] \subseteq \mathbb{B}(0;\varepsilon)$ for all $n \in \mathbb{Z}_{\geq 0}$. Finally, we define the coder-estimator scheme's *data-rate* as $R(\mathscr{S}) \coloneqq \log(\#\mathcal{A})$. ▲

The ε-accuracy conditions corresponds to the set of possible estimation errors at time n after receiving the symbols $a^{0:n-1} \in \mathcal{A}^n$. Thus, ε-accuracy means that the estimation error must be contained in a ball of radius ε for all times. Now, we can finally formally state our problem.

Problem 1
Under Assumption 1. Let $\varepsilon \in (0,\infty)$ be a prescribed estimation accuracy and $\gamma \in (0,\infty)$ be a disturbance level. What conditions we must impose on ε and γ so that an ε-accurate coder-estimator scheme (Def. 2) that operates with a finite data-rate exists for $(\mathbf{X}_n)_{n\in\mathbb{Z}_{\geq 0}}$? Under such conditions, what is the data-rate $R \in [0,\infty)$ so that i) there are no ε-accurate coder-estimator schemes with $R(\mathscr{S}) < R$ and ii) there are schemes with $R(\mathscr{S}) > R$? ▲

We call the quantity R the *minimum data-rate* for solving Problem 1. The goal of next section is to provide a partial answer to the above problem and analyze what conclusions we can draw from our results.

IV. Towards a new data-rate theorem

In this section, we present our main results, which provide a partial answer to the questions posed in Section II and Problem 1. First, we state Theorem 1, which tells us that there exists an ε-accurate coder-estimator scheme when $\varepsilon > \gamma$. We also provide data-rate lower and upper bounds in this scenario. Also, we show that the data-rate diverges when γ approaches ε. Further, when disturbances are not present, we show that the accuracy requirement forces the data-rate to be larger, in most cases, than the intrinsic entropy, which is what we need for stabilization and state reconstruction without noise [18], [22]. Finally, we analyze the results and illustrate them with an example.

The next quantity plays a major role in what follows: for $p \in [0,1)$, we define $H^p(A) \coloneqq \sum_{\sigma(A)>1-p}\log(\sigma(A))$ and we call $H^0(A)$ the *first intrinsic entropy* of A. Informally, the first intrinsic entropy is related to how much the uncertainty on the state of system (3) grows between times n and $n+1$. We note that $H^p(A)$ is not the same as the usual intrinsic entropy, as we discuss in the next section. Also, define the *accuracy-to-noise ratio* as $r \coloneqq \gamma/\varepsilon$. Finally, define the auxiliary quantities $k_p \coloneqq \max\{i \in [d] : \sigma_p(A) > 1-p\}$ and $J_p \coloneqq \max\{i \in [d] : \sigma_p(A) > \frac{1-p}{\sqrt{d}}\}$ for $p \in [0,1)$. With these additional notations, we state our main theorem.

Theorem 1
Let $\varepsilon \in (0,\infty)$ be an accuracy and $\gamma \in [0,\varepsilon)$ be the radius of the disturbance set. Under Assumption 1, there exists an ε-accurate coder-estimator scheme $\mathscr{S}$ (Def. 2) that operates with a finite data-rate that is upper-bounded by

$$R(\mathscr{S}) \leq H^r(A) + k_r \log\big(\frac{1}{1-r}\big) + \log(1.5\omega_{k_r}k_r^{k_r/2}) + O(\max\{k_r, J_r - k_r\}\log(d)) \tag{4}$$

Conversely, any coder-estimator scheme $\mathscr{S}$ that is ε-accurate for $(\mathbf{X}_n)_{n\in\mathbb{Z}_{\geq 0}}$ has its data-rate lower-bounded by

$$H^r(A) + k_r \log\big(\frac{1}{1-r}\big) \leq R(\mathscr{S}). \tag{5}$$

Finally, if $r \geq 1$, there exists no state estimator that solves the problem with finite data-rate. ▲

The expression for the term $O(\max\{k_r, J_r - k_r\}\log(d)) = k_r\log(3/2) + \frac{k_r}{2}\log(d/k_r) + (J_r - k_r)\log(2\lceil\frac{\sqrt{d}+1}{2}\rceil)$. We remark that $\log(\omega_{k_r}k_r^{k/2}) \leq k\log(\sqrt{2e\pi})$ for all $k \in \mathbb{Z}_{>0}$. The proof of this theorem is in the appendix of [**?**]. The idea of the proof of the lower bound is to perform a k_r-dimensional volume counting inside the ambient space $\mathbb{R}^d$ using concepts from multi-linear algebra, such as compound matrices (see, e.g., [2]) and geometric analysis, via the area formula (see, e.g., Theorem 5.1.1 in Section 1.1 of Chapter 5 from [12]). The proof of the upper bound involves designing an ε-accurate code-estimator scheme.

Remark 1 (Finiteness of the gap)
Note that, when $k_r = J_r = d$, inequality (4) becomes

$$R(\mathscr{S}) \leq H^r(A) + k_r \log\big(\frac{1}{1-r}\big) + k_r\log(1.5\sqrt{2e\pi}). \tag{6}$$

Comparing this to (5), when $k_r = d$, we see that the the gap between the bounds in Theorem 1 is smaller than or equal to $d\log(1.5\sqrt{2e\pi}) \approx 2.65d$ bits. Thus, the number of extra bits per dimension is constant in this case. Note that $k_r = J_r = d$ happens, for example, when the system is expansive, i.e., $\sigma_d > 1$, or when the accuracy-to-disturbance ratio r is "large". ▲

V. Analysis and Comparisons

In this section, we analyze and compare our results with the literature. We explain how our result differs from those results and what is the novelty in our analysis. Also, we revisit Example 1 to show that, when noise is present, there are preferential coordinate systems for encoding the state.

A. Comparison with the literature

Our result also shares some resemblance with the works considering stabilization of linear systems with disturbances [18]–[20]. To see why, we briefly recall that, using the uncertain variables notation, those results give a bound of the form $\lim_{n\to\infty}\sup_{m\in[0:n]}\{|x| : x \in [\![\mathbf{X}_m]\!]\} \geq \frac{\gamma}{1-2^{-(R-H(A))/d}}$, where $H(A) := \sum_{\lambda \text{ eigenvalue of } A} \max\{\log(|\lambda|), 0\}$ is the *intrinsic entropy* of A, R is the controller data-rate, i.e., the logarithm of the number symbols in the alphabet, and d is the dimension of the unstable subspace of A. Imposing the constraint $\lim_{n\to\infty}\sup_{m\in[0:n]}\{|x| : x \in [\![\mathbf{X}_m]\!]\} \leq \varepsilon$, we conclude that $R \geq H(A) + d\log(1/(1-r))$ after some algebraic steps. Note that the bound presented in that work is only asymptotic, i.e., it does not impose finite time constraints on the norm of $\mathbf{X}_n$. However, comparing $R \geq H(A) + d\log(1/(1-r))$ with (5), when $k_r = d$, we conclude that (5) is tighter.[1] Remarkably, when A is a normal matrix and noise is not considered, we have that $H(A) = H^0(A)$ since normal matrices have the property that the absolute values of their eigenvalues equal their singular value (see, e.g. [10]). We dedicate the next subsection to discuss what is behind this latter observation.

B. Coordinate-dependence

One important feature of our result that differs considerably from the usual data-rate theorems is its clear coordinate-dependence. The usual data-rate theorem involves eigenvalues, which are invariant under linear coordinate changes, whereas our results involve singular values, which vary unless the linear coordinate change is also orthogonal. The reason for this discrepancy is due to three facts: the accuracy requirement, the presence of disturbances, and the fact that transients are coordinate-dependent. First, in most practical applications, we require accuracy in a specific coordinate system, for example, maintaining a system's position and velocity, quantities with physical meaning, within a certain range. Second, the presence of disturbances together with the accuracy requirement constrains the maximum time to transmit. This fact indicates that we must understand the finite-time behavior of the dynamics, i.e., we must study the transients. Additionally, we present an example in the report [**?**] to make it clear why transients are coordinate-dependent. Interestingly, when noise is not present, we can wait to transmit and get a minimal data-rate that is as close as desired to the topological entropy, an asymptotic coordinate-invariant property of the system. Finally, we remark that, for all matrices that are diagonalizable over $\mathbb{C}$, there exists a preferential coordinate system where the minimal data-rate to solve problem 1 is the least[2].

Example 2 (Revisiting Example 1)
Consider the matrix A_1 from Example 1 and note that the eigenvalues of A_1 and A_2 are 0.9 and 0.5, which makes their intrinsic entropy equal to zero. However, A_1's singular values are $\sigma_1 \approx 2.24$ and $\sigma_2 \approx 0.2$. Hence, $H_1(A_1) \approx 1.164$ bits/sample. On the other hand, matrix A_2 is a normal matrix, which implies that its singular values equal the absolute value of its eigenvalues. Thus, $H_1(A_2) = 0$. This shows that the minimum data-rate is coordinate dependent. In Figure 1, we plot the upper and lower bounds for the the minimum data-rate when $A = A_1$ as a function of $r = \gamma/\varepsilon$. We can see that the lower bound is an increasing function of r. Further, the lower bound is close to $H_1(A_1)$ when $r \approx 0$ and both bounds grow when $r \approx 1$ as expected. The latter fact happens because we need the encoder to use finer quantization regions to compensate the presence of noise and guarantee the estimation error bound. Finally, we note that the jumps we observe in the upper bound are due to the term multiplied by $J_r - k_r$, which stars as zero, becomes one, then equals zero again. ▲

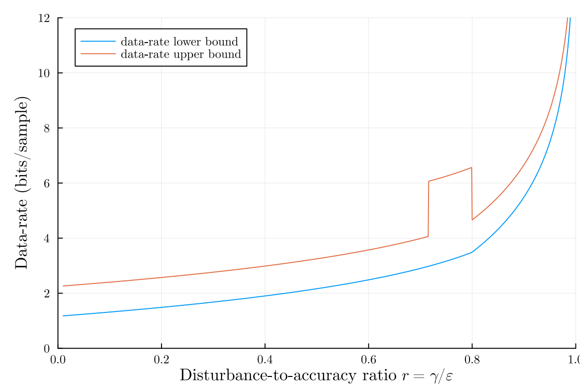

Fig. 1. Upper and lower bounds for the minimum data-rate as a function of γ/ε when $A = A_1$

VI. Conclusion

In this work, we stated the problem of reconstructing the state of a perturbed linear system from encoded measurements with a prescribed accuracy when the information is transmitted using a finite data-rate. We explained how to model the quantities involved in this problem using uncertain variables and how we can formalize tasks of this kind. Next, we explained how the minimum data-rate varies with the system dynamics, estimation accuracy, and the size of the state noise. We also showed that noise might force us to transmit data faster, unlike when no disturbances exist.

In the future, our goal will be to extend this analysis to consider transmission periods different from one. Additionally, we aim to extend this work to input-output models and derive bounds for this scenario. These results can be compared with

[1] We refer to [**?**] for a proof.

[2] See [**?**].

those obtained in [25], [30]. Furthermore, we believe that this analysis might extend other results from nonstochastic control theory, as presented in [17] and related to [14], [26].

REFERENCES

[1] J. Baillieul. *Feedback Designs in Information-Based Control*, page 35–57. Springer Berlin Heidelberg, 2002.

[2] Eyal Bar-Shalom, Omri Dalin, and Michael Margaliot. Compound matrices in systems and control theory: a tutorial. *Mathematics of Control, Signals, and Systems*, 35(3):467–521, March 2023.

[3] Dimitri Bertsekas and Ian Rhodes. Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control*, 16(2):117–128, April 1971.

[4] Franco Blanchini and Stefano Miani. *Set-Theoretic Methods in Control*. Springer International Publishing, 2015.

[5] Fritz Colonius and Christoph Kawan. Invariance entropy for control systems. *SIAM Journal on Control and Optimization*, 48(3):1701–1721, January 2009.

[6] David F. Delchamps. Extracting state information from a quantized output record. *Systems & Control Letters*, 13(5):365–372, December 1989.

[7] Gerald B Folland. *Real analysis*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. John Wiley & Sons, Nashville, TN, 2 edition, March 1999.

[8] Allen Gersho and Robert M. Gray. *Vector Quantization and Signal Compression*. Springer US, 1992.

[9] João P. Hespanha, Antonio Ortega, and Lavanya Vasudevan. Towards the control of linear systems with minimum bit-rate. In *Proc. 15th Int. Symp. on Mathematical Theory of Networks and Systems (MTNS)*, 2002.

[10] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, April 1991.

[11] Ilya Kolmanovsky and Elmer G. Gilbert. Theory and computation of disturbance invariant sets for discrete-time linear systems. *Mathematical Problems in Engineering*, 4(4):317–367, January 1998.

[12] Steven G Krantz and Harold Parks. *Geometric Integration Theory*. Cornerstones. Birkhauser Boston, Secaucus, NJ, 2008 edition, August 2008.

[13] Daniel Liberzon and Sayan Mitra. Entropy and minimal bit rates for state estimation and model detection. *IEEE Transactions on Automatic Control*, 63(10):3330–3344, October 2018.

[14] A. S. Matveev and A. V. Savkin. Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels. *International Journal of Control*, 80(2):241–255, February 2007.

[15] Alexey S. Matveev and Alexander Yu. Pogromsky. Observation of nonlinear systems via finite capacity channels, part ii: Restoration entropy and its estimates. *Automatica*, 103:189–199, May 2019.

[16] Alexey S Matveev and Andrey V Savkin. *Estimation and control over communication networks*. Control Engineering. Birkhäuser, Cambridge, MA, 2009 edition, April 2009.

[17] Girish N. Nair. A nonstochastic information theory for communication and state estimation. *IEEE Transactions on Automatic Control*, 58(6):1497–1510, June 2013.

[18] Girish N. Nair, Fabio Fagnani, Sandro Zampieri, and Robin J. Evans. Feedback control under data rate constraints: An overview. *Proceedings of the IEEE*, 95(1):108–137, January 2007.

[19] Yorie Nakahira, Quanying Liu, Terrence J. Sejnowski, and John C. Doyle. Diversity-enabled sweet spots in layered architectures and speed–accuracy trade-offs in sensorimotor control. *Proceedings of the National Academy of Sciences*, 118(22), May 2021.

[20] Yorie Nakahira, Nikolai Matni, and John C. Doyle. Hard limits on robust control over delayed and quantized communication channels with applications to sensorimotor control. In *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, December 2015.

[21] Alexander Yu. Pogromsky and Alexey S. Matveev. Remote robust state estimation for nonlinear systems. *Automatica*, 185:112795, March 2026.

[22] Andrey V. Savkin. Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control. *Automatica*, 42(1):51–62, January 2006.

[23] Fred Schweppe. *Uncertain dynamic systems*. Prentice-Hall electrical engineering series. Prentice Hall, Old Tappan, NJ, September 1973.

[24] Fred Schweppe and Harold K. Knudsen. The theory of amorphous cloud trajectory prediction. *IEEE Transactions on Information Theory*, 14(3):415–427, May 1968.

[25] Hidenori Shingin and Yoshito Ohta. Amplitude snr limitation in control over channels with bounded noise. *Automatica*, 97:286–291, November 2018.

[26] S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Transactions on Automatic Control*, 49(7):1056–1068, July 2004.

[27] Guilherme S. Vicinansa and Daniel Liberzon. Estimation entropy, lyapunov exponents, and quantizer design for switched linear systems. *SIAM Journal on Control and Optimization*, 61(1):198–224, February 2023.

[28] Hans S. Witsenhausen. Sets of possible states of linear systems given perturbed observations. *IEEE Transactions on Automatic Control*, 13(5):556–558, October 1968.

[29] Guosong Yang, A. James Schmidt, Daniel Liberzon, and João P. Hespanha. Topological entropy of switched linear systems: general matrices and matrices with commutation relations. *Mathematics of Control, Signals, and Systems*, 32(3):411–453, September 2020.

[30] Keyou You, Lihua Xie, Shuli Sun, and Wendong Xiao. Quantized filtering of linear stochastic systems. *Transactions of the Institute of Measurement and Control*, 33(6):683–698, August 2011.