A Tutorial on Quantizer Design for Networked Control Systems: Stabilization and Optimization

Serdar Yüksel[1]

Abstract

In this article, a tutorial on optimal quantizer design in networked control systems is presented. The goal of this invited article is not to present a literature survey (although an effort is presented for a brief literature review); but to present a detailed discussion on the salient issues and some of the main results in the literature and by the author and collaborators, together with some previously unpublished contributions. The analysis is restricted to noiseless channels with finite capacity. The analysis and results presented consist of quantizer design for stabilization over noiseless channels, optimal quantizer design for minimizing a class of cost functions, quantizer design for stabilization of decentralized systems, and optimal quantizer design under a class of decentralized information structures.

1 Networked Control Systems

A Networked Control System is a control system in which the components are connected through real-time communication channels or a data network. Thus, there may be a data link between the sensors (which collect information), controllers (which make decisions), and actuators (which apply the controller commands).

Due to such a networked structure, many modern control systems are decentralized. A system is said to be decentralized if there are multiple decision makers in the system (e.g., sensors, controllers, encoders) and these decision makers have access to different and imperfect information with regard to the system they operate in, and they need to either cooperate or compete with each other. Such systems are becoming ubiquitous, with applications ranging from automobile and inter-vehicle communications design, control of surveillance and rescue robot teams for access to hazardous environments, space exploration and aircraft design, among many other fields of applications (see [73], [26], [43] for a review of application areas).

In such remote control applications, one major concern is the characterization of a sufficient amount of information transfer needed for a satisfactory performance. This information transfer can be between various components of a networked control system. One necessity for satisfactory control performance is the ability for the controllers to track the plant state under communication constraints. One other challenge is the determination of the data rate required for the transmission of control signals, and the construction of dynamic encoding, decoding, and control policies meeting some criteria. Another important problem is the coordination among multiple sensors or multiple controllers/decision makers with the lowest information exchange possible. Even in cases when communication resources are not scarce, a strong understanding of the fundamentals can be useful in the system architecture, and finally, such an insight can help reduce the computation requirements and complexity.

As indicated earlier, there are usually three types of agents involved in a control system: sensors, controllers, and actuators (plants) (see Fig. 1). Various forms of this architecture have been introduced and studied in the literature in the networked control community. To be able to analyze different scenarios, it is important to identify the probabilistic description of the system and characterize the information structure in the system. Toward this goal; in the following we review information structures in decentralized control.

[1]S. Yüksel is with the Department of Mathematics and Statistics at Queen's University, Kingston, Ontario, Canada, K7L 3N6. Email: yuksel@mast.queensu.ca.

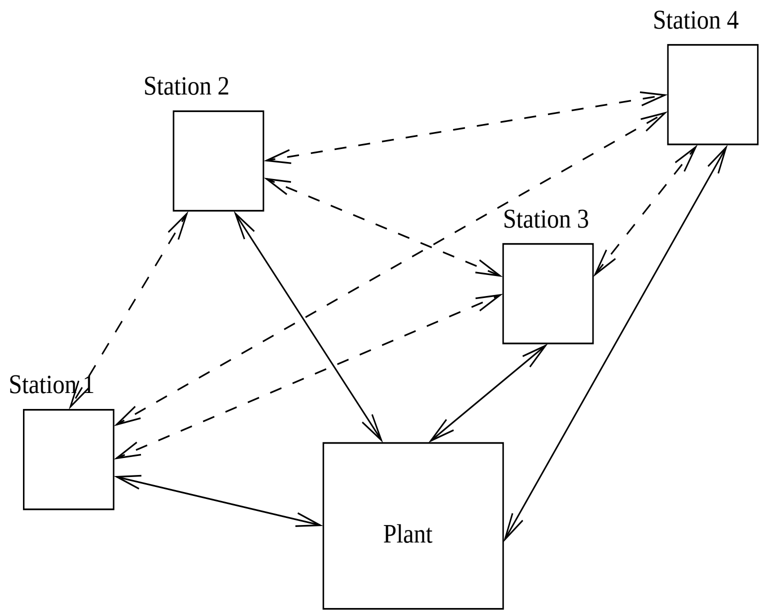

Figure 1: A decentralized networked control system. Solid lines show the interaction between the control stations and the plant. Dashed lines depict the possible communication links between the stations.

1.1 Information Structures in a Networked Control System

In the following, we present a general discussion on information structures; which will also be useful in our analysis for quantized systems.

We can present a general decentralized system based exclusively on the underlying probability space and the interaction dynamics of the decentralized decision makers. Such a model was considered by Witsenhausen, and is known as Witsenhausen's Intrinsic Model [153]. In this model, any action applied at a given time stage is regarded as applied by an individual station/decision maker/agent. The Intrinsic Model of Witsenhausen has three components:

1. An information structure: $\mathcal{I} := \{\Omega, \mathcal{F}, (U^k,\mathcal{U}^k), 1 \leq k \leq N\}$ specifying the system's allowable decisions and events. N is the number of control actions taken, and each of these actions is taken by an individual station (hence, even a station with perfect recall can be regarded as a separate decision maker, at different time stages). $(\Omega, \mathcal{F})$ is the measurable space generating the probability space. $(U^k,\mathcal{U}^k)$ denotes the measurable space from which the action of decision maker k, u^k is selected.

2. A design constraint, which restricts the set of admissible N-tuples, control laws $\gamma = \{\gamma^1, \gamma^2, \dots, \gamma^N\}$, called designs to the set of all measurable $\mathcal{I}^k$ measurable control functions. Here, the sub-sigma field $\mathcal{I}^k$ denotes the information that can be used to select the kth control action.

3. A probability measure P defined on $(\Omega, \mathcal{F})$.

A decentralized problem is dynamic if the information available at one decision maker (DM) is affected by the actions of another decision maker. A decentralized problem is called static, if the information available at every decision maker is only affected by exogenous disturbances, or nature; that is no other decision maker can affect the information at any given decision maker.

Witsenhausen defines information structures to be one of the three types: Classical, quasi-classical and non-classical. If the information available at the decision makers is nested in the sense that $\mathcal{I}^k \subset \mathcal{I}^{k+1}$, the information structure is *classical*. An information structure is *quasi-classical*, if whenever the control actions of a DM^i affects the observations of another decision maker DM^j, the information available at DM^i is known noiselessly by the affected decision maker. An information structure is non-classical, if it is not quasi-classical.

In a large class of applications, there is a pre-defined, deterministic, ordering of control actions, and there is a *state-space* model; in which there are a number of decision makers with perfect recall (that is, these decision makers remember their past information, with growing time, the information fields are expanding), acting sequentially; we call this model a sequential model, following Witsenhausen.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathbb{X}$ be a complete, separable, metric space in which realizations of a random sequence $\{x_t, t \in \mathbb{Z}_+ \cup \{0\}\}$ measurable on the probability space live in. Let $\mathbb{Y}^i$, be another space for $i = 1, 2, \ldots, L$ and let an observation channel $\mathcal{C}^i$ be defined as a stochastic kernel (regular conditional probability measure) on $\mathbb{X} \times \mathbb{Y}^i$, such that for every $x \in \mathbb{X}$, $P(.|x)$ is a probability distribution on the (Borel) sigma-algebra $\mathcal{B}(\mathbb{Y}^i)$ and for every $A \in \mathcal{B}(\mathbb{Y}^i)$, $P(A|.)$ is a measurable function of x.

Let there be L decision makers, $\{\text{DM}^i, i = 1, 2, \ldots, L\}$. Let a Decision Maker (DM) DM^i be located at one end of an observation channel $\mathcal{C}^i$, with inputs x_t generated as $y_t^i \in \mathbb{Y}^i$ at the channel output. We refer to a policy Π^i as a sequence of control functions measurable with respect to the sigma-algebra generated by

$$I_t^i = \{y_t^i, z_t^i; y_{[0,t-1]}^i, u_{[0,t-1]}^i, z_{[0,t-1]}^i\} \quad t \geq 1,$$

$$I_0^i = \{y_0^i, z_0^i\},$$

with control actions $u_t^i \in \mathbb{U}^i$, with the notation for $t \geq 1$

$$y_{[0,t-1]}^i = \{y_s^i, 0 \leq s \leq t-1\}.$$

Here z_t^i denotes the additional information that can be supplied to DM^i at time t.

Let DM^i have a policy Π^i and under this policy generate control actions $\{u_t^i, t \geq 0\}$, $u_t^i \in \mathbb{U}^i$, and let a dynamical system and observation channels be described by the following discrete-time equations:

$$x_{t+1} = f(x_t, u_t^1, u_t^2, \ldots, u_t^L, w_t),$$

$$y_t^i = g^i(x_t, v_t^i),$$

for some measurable functions $f, \{g^i\}$, with $\{w_t\}$ independent, identical, white system noise process and $\{v_t^i, i = 1, 2, \ldots, L\}$ be disturbance processes. Let $\mathbb{X}^T = \prod_{t=0}^{T-1} \mathbb{X}$ be the $T-$product space of $\mathbb{X}$. For the above setup, under a sequence of control policies $\{\Pi^1, \Pi^2, \ldots, \Pi^L\}$, we define an *Information-Control Structure* (ICS) as a probability space

$$(\mathbb{X}^T \times \prod_{t=0}^{T-1} \prod_{k=1}^{L} \mathbb{Y}^k \times \prod_{t=0}^{T-1} \prod_{k=1}^{L} \mathbb{U}^k, \mathcal{B}(\mathbb{X}^T \times \prod_{i=0}^{T-1} \prod_{k=1}^{L} \mathbb{Y}^k \times \prod_{i=0}^{T-1} \prod_{k=1}^{L} \mathbb{U}^k), P)$$

Here, P is the probability measure on the (Borel) sigma-algebra $\mathcal{B}(\mathbb{X}^T \times \prod_{i=0}^{T-1} \prod_{k=1}^{L} \mathbb{Y}^k \times \prod_{i=0}^{T-1} \prod_{k=1}^{L} \mathbb{U}^k)$.

*Information Pattern*s determine the sub-fields for all decision makers and time stages $\sigma(I_t^i) \subset \mathcal{B}(\mathbb{X}^T \times \prod_{i=0}^{T-1} \prod_{k=1}^{L} \mathbb{Y}^k \times \prod_{i=0}^{T-1} \prod_{k=1}^{L} \mathbb{U}^k)$. Hence, the control policies are measurable on the sub-fields, which are characterized by I_t^i for all DMs, through the term z_t^i. Thus, an Information Pattern determines what the control action can depend on, inducing an information-control structure. With the above formulation, in general the objective of the decision makers is the minimization:

$$E_{x_0}^{\Pi^1, \Pi^2, , \ldots, \Pi^L}[\sum_{i=0}^{T-1} c(x_t, u_t^1, u_t^2, \ldots, u_t^L)],$$

over all policies $\Pi^1, \Pi^2, \ldots, \Pi^L$, with initial condition x_0. Here, $E_{x_0}^{\Pi^1, \Pi^2, , \ldots, \Pi^L}[]$ denotes the expectation over all sample paths with initial state given by x_0 under policies $\{\Pi^1, \Pi^2, \ldots, \Pi^L\}$.

For a general vector q, let $\mathbf{q}$ denote $\{q^1, q^2 \ldots, q^L\}$. Let $\mathbf{\Pi} = \{\Pi^1, \Pi^2, \ldots, \Pi^L\}$ denote the ensemble of policies. Under an ensemble of policies $\mathbf{\Pi}$ and a given information pattern, with an initial condition x_0, the attained performance index is

$$J_{x_0}(\mathbf{\Pi}) = E^{\mathbf{\Pi}}_{x_0}[\sum_{t=0}^{T-1} c(x_t, \mathbf{u}_t)]$$

We say the information available at DM^i is nested in that of DM^j at time t, if $\sigma(I^i_t) \subset \sigma(I^j_t)$. Nestedness has very important implications; but as indicated in [172] such a characterization of information sets is too strong. It was observed by Radner [121] that a static LQG team problem with a non-nested information structure admits an optimal solution which is linear.

An information structure is *partially nested*, if whenever the control actions of a DM^i affects the observations of another decision maker DM^j, the information available at DM^i is known noiselessly by the affected decision maker, that is: $Z^j_t = \{y^i_t, \quad \text{if} \quad \mathrm{DM^i} \rightarrow \mathrm{DM^j}\}$. Here the notation DM^i $\rightarrow$ DM^j denotes the fact that the actions of DM^i affects the information at DM^j (which is also known as signaling, see [176] for a review of signaling). The partially nested structure effectively reduces the dynamic LQG team problem to a static optimization problem in the sense that the signaling (inner) agent (whose information sigma algebra is a subset of the signaled (outer) agent's information sigma algebra) makes all her decisions statically and the outer agent can generate such *pure strategy* decisions and the joint decisions can be regarded as one single-DM's decision, effectively making the problem static among such single DMs. Partially nested structures can also have a dynamic evolution [27], and as a special case, this includes the case where information propagation is faster than dynamics propagation, where in the above definition, delay is also considered [27], [147], [122]. According to Witsenhausen's definition, quasi-classical information structure is equivalent to partial nestedness. This structure includes the *one-step delayed observation sharing* information pattern (see [80] and [9]), which allows the Decision Makers to share all their observations with a unit delay: $z^i_t = \{\mathbf{y}_{t-1}\}$. If the agents also share their decisions, then the information pattern is called *one-step delayed information sharing* pattern: $z^i_t = \{\mathbf{y}_{t-1}, \mathbf{u}_{t-1}\}$. For further related discussion, please see [163].

If a decision maker's, DM^j, information is dependent on the actions of another, say DM^k, and DM^j does not have access to the information available to DM^k, this information structure is said to be *non-classical*. Hence, an information pattern which is not partially nested is a non-classical information pattern. The *one-step delayed control sharing pattern* $z^i_t = \mathbf{u}_{t-1}$ is one such example [4], [129], [9].

A systematic approach for generating optimal control policies by eliminating redundant information has been introduced in [88].

[147], [6] and [123] have studied sufficiency conditions for tractability and convexity in optimal decentralized control problems.

Even when the information structure is non-classical, one may obtain tractable solutions for optimal control problems. The *stochastically nested information structure* is an example of such problems [172]. Shannon's point-to-point communication problem can be regarded as another example which is by definition non-classical, but tractable.

When the information structures are non-nested, controllers might choose to communicate via their control actions, that is might wish to pursue *signaling*. Different types of signaling can occur: signaling to learn the dynamics of the system, signaling what the belief (that is, the conditional probability measure) on the state of the system is, signaling what the belief on the other agents controls are or signaling what the agent's own future control actions will be, depending on the effects on the cost performance. Especially, in decentralized control systems, the notion of information, and how it is generated, is often part of the control problem itself, and this has been captured in the so-called *dual effect* [8], so that control can serve both to improve the quality of estimation and also to achieve its primary objective. This leads to a *triple effect*, when there is also an incentive for signaling: the action that decision makers communicate each other through the plant. This arises prominently in networked control systems. We conclude this section by noting that, the information structures can be further identified, by also considering the probability measure and the cost functions; such an approach has been adopted in [91].

1.2 An Incomplete Literature Review on Quantizer Design in a Class of Networked Control Systems

In the following, we present an incomplete literature review on the fundamental bounds for networked control systems, where we restrict the analysis to noiseless channels with finite capacity.

We refer the reader to [106] for a literature survey of quantizer design in networked control problems and the comprehensive book [99] for a detailed analysis of the literature as well as more general noisy settings. The goal here is not to present a survey; such a task of writing a genuine survey is beyond what we could provide.

As mentioned earlier, the use of digital and wireless channels such as the Internet or bus lines (as in a Controller Area Network (CAN)) in control systems has become common place raising important mathematical (control theoretic and information theoretic) challenges for design of such systems. The design of such systems combining ideas from information theory, control theory and applied probability had already been investigated in 1950s to 1970s, for example in the context of state space design and Kalman Filtering [75], estimation [39], design of statistical experiments [14], and value of information channels [35].

There is a large literature on stochastic stabilization of sources via coding, both in the information theory and control theory communities. In the information theory literature, stochastic stability results are established mostly for stationary sources, which are already in some appropriate sense stable sources. In this literature, the stability of the estimation errors as well as the encoder state processes are studied. These systems mainly involve causal and non-causal coding (block coding, as well as sliding-block coding) of stationary sources [76], [51], and asymptotically mean stationary sources [55]. Real-time settings such as sigma-delta quantization schemes have also been considered in the literature, see for example [160] among others. Earlier papers in control theory literature on quantized estimation include [52].

There also have been important contributions on non-causal coding of non-stationary/unstable sources: Consider the following Gaussian AR process:

$$x_t = -\sum_{k=1}^{m} a_k x_{t-k} + w_t,$$

where $\{w_k\}$ is an independent and identical, zero-mean, Gaussian random sequence with variance $E[w_1^2] = \sigma^2$. If the roots of the polynomial: $H(z) = 1 + \sum_{k=1}^{m} a_k z^{-k}$ are all in the interior of the unit circle, then the process is stationary and its rate distortion function (with the distortion being the expected, normalized Euclidean error) is given parametrically by the following [56], obtained by considering the asymptotic distribution of the eigenvalues of the correlation matrix:

$$D_\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} \min(\theta, \frac{1}{g(w)})dw,$$

$$R(D_\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \max(1/2(\log\frac{1}{\theta g(w)}), 0)dw,$$

with $g(w) = \frac{1}{\sigma^2}|1 + \sum_{k=1} a_k e^{-ikw}|^2$. If at least one root, however, is on or outside the unit circle, the analysis is more involved as the asymptotic eigenvalue distribution contains unbounded components. [56] and [54] showed that, using the properties of the eigenvalues as well as Jensen's formula for integrations along the unit circle, $R(D_\theta)$ above should be replaced with:

$$R(D_\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \max(\frac{1}{2}\log(\frac{1}{\theta g(w)}), 0)dw + \sum_{k=1}^{m} \frac{1}{2}\log(|\rho_k|^2), \quad (1)$$

where $\{\rho_k\}$ are the roots of the polynomial.

[16] obtained the rate-distortion function for Wiener processes, and in addition, developed a two-part coding scheme, which was later generalized in [127], to unstable Markov processes driven by bounded noise. The scheme in [16] exploits the independent increment property of Wiener processes.

Thus, an important finding in the above literature is that, the logarithms of the unstable poles in such linear systems appear in the rate-distortion formulations, an issue which has also been observed in the networked control literature, which we will discuss further below. We also wish to emphasize that these coding schemes are non-causal, that is the encoder has access to the entire ensemble before the encoding begins, or the coding is a sliding-block/sliding-window scheme with a finite degree of non-causality.

In contrast with information theory, due to the practical motivation of sensitivity to delay, the control theory literature has mainly considered causal/zero-delay coding for unstable (or non-stationary) sources, in the context of networked control systems.

There has been a renewed and growing interest with the emergence of practical applications within the past decade. In this context, one of the earliest papers on the topic is [7], which has shown that for a scalar discrete-time linear Gaussian system controlled over a Gaussian channel, the encoder and the controllers with noiseless causal feedback which jointly minimize a quadratic objective functional are all linear. This was perhaps the first paper that used information theory along with stochastic control in the analysis of a control system. References [155] and [149] studied the optimal causal coding problem over respectively a noiseless channel and a noisy channel with noiseless feedback. Reference [37] showed the chaotic nature of quantization in control in one of the first papers to bring in quantization as a design limitation; [22], on the other hand, studied the trade-off between delay and reliability, and formulated relevant and challenging problems; the questions that were posed there led to an accelerated pace of research efforts on the topic: Significant progress on the connection between information theory and control has been achieved through study of the minimum information rate requirements needed for stabilizability over noisy channels with noiseless feedback, under various assumptions on the noise models, as well as control over noiseless channels—as reported in [161], [143], and [108], where [108] also considered a class of quantizer policies for systems driven by noise with unbounded support set for its probability measure. References [161], [143], and [108], obtained the minimum lower bound needed for stabilization over noisy channels under a class of assumptions on the system noise and channels; known as the data rate theorem. This theorem states that for stabilizability under information constraints, in the mean-square sense, a minimum rate needed for stabilizability has to be at least the sum of the logarithms of the unstable poles/eigenvalues in the system; that is:

$$\sum_{k=1}^{m} \frac{1}{2} \log(|\rho_k|^2),$$

This result could be contrasted with (1).

Several studies in the literature have focused on noiseless discrete channels with time-invariant encoders, where the main issue becomes one of design of an invariant quantizer; see [42] and [44]. The first of these, [42], adopts a Lyapunov-based approach to stabilize a system with limited information, and shows that the coarsest quantizer achieving stability is logarithmic and that the design is universal, i.e., it has the same base for construction regardless of the sampling interval. The second reference, [44], quantifies the relationship between the rate of convergence and communication rate for scalar systems, and provides conditions for set invariance as a measure of stability. An information theoretic approach to quantizer design leading to stabilizability was presented in [177]. [93] and [95] consider the disturbance rejection problem over noisy channels. In [73], a Lyapunov theoretic quantizer design leading to stability is provided, with further results reported in [72], and time-varying quantizer design presented also in [126]. Reference [29] studied control over channels with partial observations, and adopted an information theoretic approach, also with a robust formulation on the class of sources. For noisy binary forward channels with noiseless feedback, coding schemes were presented for the forward channel with noiseless feedback in [133], and linear time-invariant as well as time-varying control policies on continuous-alphabet channels modeled as Gaussian channels were considered in [24].

For control over erasure channels there are also various studies: [71], [103], [164], [180], [167], [132], [60]. For more general noisy channels, we refer the reader to [99], [93], [92], [41], [127], [181], among others.

For coding and information transmission for unstable linear systems, there is an important difference between continuous alphabet and finite-alphabet (discrete) channels as discussed in [181]: When the space is continuous alphabet, we do not necessarily need to consider adaptation in the encoders. On the other hand, when the channel is finite alphabet, and the system is driven by unbounded noise, a bounded range quantizer (a quantizer with bounded granular region) leads to almost sure instability. This was first recognized in view of the unboundedness of second moments in Proposition 5.1 in [108], transience of the process is established in Theorem 4.2 in [181]. [30] considers the conditions for stabilization when the control signals are bounded.

As discussed above, zooming type adaptive quantizers, introduced by Brockett and Liberzon [25], for remote stabilization of open-loop unstable, noise-free systems with arbitrary initial conditions has been very useful. Reference [171] obtained a martingale characterization for stabilization for such zooming quantizers even when there is noise in the system, and the noise has unbounded support for its probability measure. This approach has been extended to control over erasure channels in [167] and [180].

There is also a large body of literature on real-time quantizer design in the communications and information theory community, as conveniently presented in the survey paper [57]. There are many important contributions from this field on optimal quantization of general sources; in both in the context of block-codes (where infinite copies of a source are encoded simultaneously), as well as in the single-shot, or delay-limited settings. In this context, we wish to point out one observation: The distortion-rate problem [34] and the entropy-constrained quantization problem [61] exhibit the distinctness of the two problems: The distortion-rate (or the dual problem of rate-distortion) deals with efficient coding of an infinite copy of a source; whereas the entropy-constrained quantization problem deals with a real-time setup of encoding one realization of a random variable.

One important reference is the work by Goodman and Gersho [51], where an adaptive quantizer was introduced and the adaptive quantizer's stationarity properties were investigated when the source fed to the quantizer is a second order and i.i.d. sequence. In fact, zooming type quantizers is a special class of Goodman and Gersho's adaptive quantization scheme. Kieffer and Dunham [76], have obtained conditions for the stochastic stability of a number of coding schemes when the source considered is also stable, where various forms of stability of the quantizer and the estimation error have been studied.

There is also a large collection of work in the context of optimal real-time source coding, Related papers on real-time coding include the following: [111] established that the optimal optimal causal encoder minimizing the data rate subject to a distortion for an i.i.d sequence is memoryless. If the source is kth-order Markov, then the optimal causal fixed-rate coder minimizing any measurable distortion uses only the last k source symbols, together with the current state at the receiver's memory [155]. References [149] considered the optimal causal coding problem of finite-state Markov sources over noisy channels with feedback. [145] and [89] considered optimal causal coding of Markov sources over noisy channels without feedback. [87] considered the optimal causal coding over a noisy channel with noisy feedback. Reference [83] considered the causal coding stationary sources under high-rate assumption. Borkar, Mitter and Tatikonda [19] studies a related problem of coding of a partially observed Markov source, however, the construction for the encoders is restricted to take a particular form which uses the information at the decoder and the most recent observation at the encoder (not including the observation history). [110] considers decentralized coding of correlated sources when the encoders observe conditionally independent messages given a finitely valued random variable and obtain separation results for the optimal encoders. The paper also considers noisy channels. Reference [169] and [170] consider partially observed sources, and some results in these papers will be presented in this article. References [151] and [104] consider optimal causal variable-rate coding under side information and [178] considers optimal variable-rate causal coding under distortion constraints. A parallel line of consideration which has a rate-distortion theoretic nature is on *sequential-rate distortion* proposed in [143] and the *feedforward* setup, which has been investigated in [148] and [40]. A related work is Witsenhausen's indirect rate distortion problem [154] (see also [38]). Further related papers include [12], [69].

This paper also presents a discussion for the decentralized setup. The corresponding brief literature review will be presented in Section 5.

Here is a brief overview of the rest of the paper. In the next section, we revisit some preliminary definitions on quantization and information theoretic notions. In Section 3, we consider the problem of stabilization under quantization constraints, and in Section 4, we consider the problem of optimal singleterminal quantization. In Section 5, stabilization of decentralized systems under quantization constraints is discussed. In Section 6, optimization for decentralized systems is considered. The paper ends with the concluding remarks of Section 7

2 Quantization and Preliminaries

Essential in communications problems is the embedding of information into a finite set, possibly with loss of information. This is done through quantization, which is a mapping from a larger alphabet to a smaller alphabet. We refer the reader to the survey paper [57] on a detailed review of quantization and introduction of different types of quantizers.

Let us present a formal definition which we will adopt throughout the paper.

Definition 2.1 *Let $\mathcal{M} = \{1, 2, \ldots, M\}$ with $M = |\mathcal{M}|$. Let $\mathbb{A}$ be a topological space. A quantizer $Q(\mathbb{A}; \mathcal{M})$ is a Borel measurable map from $\mathbb{A}$ to $\mathcal{M}$.* ⋄

We define the bins or cells in a quantizer as the sets:

$$\mathcal{B}_i = \{a \in \mathbb{A} : Q(a) = i\}, \quad i \in \mathcal{M}.$$

We note that, traditionally, in source coding theory, a quantizer is also characterized by a reconstruction value in addition to a set of partitions; in our definition we only define the quantizer by the bins in the quantizer.

We say the *rate* of such a quantizer is $\log_2(|\mathcal{M}|)$ bits.

When the spaces $\mathbb{A}$ and $\mathcal{M}$ are clear from context, we will drop the notation and denote the quantizer simply by Q.

In practice, however, one may impose additional structure on the quantizers; for example one may impose the bins in the quantizers to be convex. Further structures are possible in this context. An important class of quantizers on $\mathbb{R}^n$ are those which are uniform, those which have their bins with identical Lebesgue measure. These will be discussed later in the paper.

The information rate is measured by the number of bits. When the coding is variable-rate, information is measured by the average number of bits needed to be exchanged among the decision makers; whereas when the coding scheme is fixed-rate, information is measured by the actual number of bits that are exchanged for any given time stage $t \geq 0$.

Let x be a finitely-valued random variable. Consider the following problem: What is the minimum amount of average bits needed to represent x with no loss? The answer to this leads us to the notion of *entropy* of a source [34]: Entropy of a finitely-valued random variable provides an almost tight lower bound on the average bit rate needed to compress the source losslessly. The entropy of such a source is defined as $H(x) = -\sum p(x) \log_2(p(x))$. When the source is continuous, we use the term *differential entropy*: $h(x) = -\int p(dx) \log_2(p(dx))$, where $p(dx)$ is the measure of the source.

Mutual information between an input random variable, x, and a corresponding output, y, is

$$I(x; y) = H(x) - H(x|y)\,,$$

where $H(x|y)$ is the conditional entropy of x given y.

We would like to point out one important distinction between information theoretic applications and such real-time problems considered in the paper. The relation between the rate-distortion function and the notion of *distortion-constrained entropy minimization* [61] [62] [177] is a striking example identifying the difference: In rate-distortion theoretic analysis, an infinite realization of a random variable is observed, and compression is performed based on the entire sequence. For a quantization problem; however, a single realization is observed, and the output it generated only for the single realization. Clearly, the rate-distortion function leads to a lower value, since one could *shape* the codebooks appropriately [83]. Only for quantization of vector valued processes, in the limit of large dimensions, one observes an equivalence between the distortion-constrained entropy minimization and the rate-distortion function [57]. We note that, in the definition of the rate distortion function $R(D)$ for a random variable x taking realizations in $\mathbb{X}$ with distortion metric $\rho : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$:

$$R(D) = \inf_{P(\hat{x}|x), \gamma : E[\rho(x,\gamma(\hat{x}))] \leq D} I(x;\hat{x}),$$

the minimization is over the space of stochastic kernels $P(\hat{x}|x)$, where $I(\cdot,\cdot)$ denotes the mutual information function. In the quantization framework, the optimization is over the space of quantizer-decoder pairs only. Note that, when $\hat{x}$ admits a discrete probability measure and $\hat{x} = \gamma(Q(x))$, for some decoder function γ, $I(x;\hat{x})$ reduces to $H(\hat{x})$. Thus, the essential difference is the space of optimization. Furthermore, even though the rate-distortion function is a convex function of distortion, the distortion-constrained entropy function is not a convex function of distortion (see [61], [62], where in the former study [61], an analysis for a uniform source has been considered). The space of quantizers is not a convex space.

Although typical information-theoretic approaches require long delays; important special cases provide optimal performance even under delay-limited settings: One example is the problem of communicating a Gaussian source over a Gaussian channel. This topic is related to the matching principle; the readers are referred to [49] and [7]. On a parallel note, the multi-terminal source coding theorems [34], although insightful, are not always applicable for a real-time setting, as the asymptotic partitioning arguments in classical information theory [34] do not apply. In a control context, however, one method to achieve the information theoretic bounds is via binning; see [175], [176], [58] for discussions on binning in a decentralized control context and [120] for a discussion on binning in a general communications context.

2.1 General Setup for a Quantization Problem: Policies, Actions, Causality and Measurability

This subsection considers a typical optimal causal encoding/quantization setup in a networked control system. For simplicity of the setup, we consider only two encoders, for a decentralized system and use this system to introduce the causality and measurability constraints in quantizer design.

We begin with providing a description of the system model. We consider a partially observed Markov process, defined on a probability space, again, $(\Omega, \mathcal{F}, P)$ and described by the following discrete-time equations for $t \geq 0$ (which may be modified to include control actions as well):

$$x_{t+1} = f(x_t, w_t), \tag{2}$$
$$y_t^i = g^i(x_t, r_t^i), \tag{3}$$

for (Borel) measurable functions $f, g^i, i = 1, 2$, with $\{w_t, r_t^i, i = 1, 2\}$ noise processes, which are independent across time and space. Here, we let $x_t \in \mathbb{X}$, and $y_t^i \in \mathbb{Y}^i$, where $\mathbb{X}, \mathbb{Y}^i$ are complete, separable, metric spaces (Polish spaces), and thus, include countable spaces or $\mathbb{R}^n$, $n \in \mathbb{Z}_+$.

Let an encoder, Encoder i, be located at one end of an observation channel characterized by (3). The encoders transmit their information to a receiver (see Figure 5), over a discrete noiseless channel with finite capacity; that is, they quantize their information.

We refer by a **Composite Quantization Policy** $\Pi^{composite,i}$ of Encoder i, a sequence of functions $\{Q_t^{composite,i}, t \geq 0\}$ which are causal such that the quantization output at time t, q_t^i, under Π^i is generated by a causally measurable function of its local information, that is, a mapping measurable with respect to the sigma-algebra generated by

$$I_t^i = \{y_{[0,t]}^i, q_{[0,t-1]}^1, z_{[0,t-1]}^1\}, \quad t \geq 1,$$

and

$$I_0^i = \{y_0^i\},$$

to $\mathcal{M}_t^i$, where

$$\mathcal{M}_t^i := \{1, 2, \ldots, |\mathcal{M}_t^i|\},$$

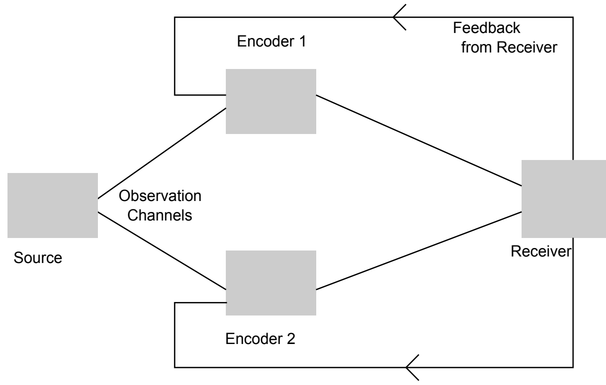

Figure 2: Partially observed source under a decentralized structure.

for $0 \leq t \leq T-1$ and $i = 1, 2$. Here z_t^i denotes some additional side information available, such as feedback from the receiver. Here, we have the notation for $t \geq 1$, as earlier:

$$y^i_{[0,t-1]} = \{y^i_s, 0 \leq s \leq t-1\}.$$

Let $\mathbf{I}^i_t$ be the space such that for all $t \geq 0$, $I^i_t \in \mathbf{I}^i_t$. Thus,

$$Q_t^{composite,i} : \mathbf{I}^i_t \to \mathcal{M}^i_t.$$

We may express, equivalently, the policy $\Pi^{composite,i}$ as a composition of a **Quantization Policy** Π^i and a **Quantizer**.

A quantization policy of encoder i, $\mathcal{T}^i$, is a sequence of functions $\{T^i_t\}$, such that for each $t \geq 0$, T^i_t is a mapping from the information space $\mathbf{I}^i_t$ to the space of quantizers $\mathbb{Q}^i_t$. A quantizer, subsequently is used to generate the quantizer output. That is for every t and i, $\Pi^i(I_t) \in \mathbb{Q}^i_t$ and for every $I^i_t \in \mathbf{I}^i_t$,

$$Q_t^{composite,i}(I^i_t) = (T^i_t(I^i_t))(I^i_t), \tag{4}$$

mapping the information space to $\mathcal{M}^i_t$ in its most general form.

Even though there may seem to be duplicated information in (4) (since a map is used to pick a quantizer, and the quantizer maps the available information to outputs) we will eliminate any informational redundancy: A quantizer action will be generated based on the common information at the encoder and the receiver, and the quantizer will map the relevant private information at the encoder to the quantization output.

Let the information at the receiver at time $t \geq 0$ be $I^r_t = \{q^1_{[0,t-1]}, q^2_{[0,t-1]}\}$. Let the common information, under feedback information, at the encoders and the receiver be the set I^c_t. Thus, we can express the composite quantization policy as:

$$Q_t^{composite,i}(I^i_t) = (T^i_t(I^c_t))(I^i_t \setminus I^c_t),$$

mapping the information space to $\mathcal{M}^i_t$.

We note that, any composite quantization policy can be expressed in the form above; that is there is no loss in the space of causal policies.

Thus, we let DM^i have policy Π^i and under this policy generate **quantizer**s $\{Q^i_t, t \geq 0\}$, $Q^i_t \in \mathbb{Q}^i_t$ (Q^i_t is the quantizer used at time t). Under action Q^i_t, the encoder generates q^i_t, as the *quantization output* at time t.

The receiver, upon receiving the information from the encoders, generates its decision at time t, also causally: An admissible causal receiver policy is a sequence of measurable functions $\gamma = \{\gamma_t\}$ such that

$$\gamma_t : \prod_{s=0}^{t} \left(\mathcal{M}^1_s \times \mathcal{M}^2_s\right) \to \mathbb{U}, \qquad t \geq 0$$

where $\mathbb{U}$ denotes the the decision space.

Let us present further notation. As denoted earlier, for a general vector a, let $\mathbf{a}$ denote $\{a^1, a^2\}$ and let $\mathbf{\Pi} = \{\Pi^1, \Pi^2\}$ denote the ensemble of policies and $\mathbf{Q}_t = \{Q_t^1, Q_t^2\}$. Hence, $\mathbf{q}_{[0,t]}$ denotes $\{q_{[0,t]}^1, q_{[0,t]}^2\}$.

With the above formulation, for example, one typical objective functional of the decision makers is the following

$$\inf_{\mathbf{\Pi}^{composite}} \inf_{\gamma} E_{\nu_0}^{\mathbf{\Pi}^{composite},\gamma}[\sum_{t=0}^{T-1} c(x_t, v_t)],$$

over all policies $\mathbf{\Pi}^{composite}$, γ with initial condition distribution ν_0. Here $c(x_t, v_t)$, is a non-negative function and $v_t = \gamma_t(\mathbf{q}_{[0,t]})$ for $t \geq 0$. This setup may include, for the case when $\mathbb{X} = \mathbb{R}^n$, the cost function: $c(x, v) = \|x - v\|_2^2$.

Another objective is to ensure that a dynamical system is stable in some appropriate sense.

We will consider these in the following sections. First, we discuss the stabilization problem.

3 Quantizer Design for Stabilization over Noiseless Channels

We consider a multi-dimensional linear system connected over a noiseless channel.

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad y_t = x_t, \tag{5}$$

where $x_t \in \mathbb{R}^n$ is the state at time t, u_t is the control input, and $\{w_t\}$ is a sequence of zero-mean independent, identically distributed (i.i.d.) $\mathbb{R}^n$-valued zero-mean Gaussian random variables. Furthermore, $E[||w_t||_2^2] = E[||w_1||_2^2] < \infty$. Here A is the system matrix with at least one eigenvalue greater than 1 in magnitude, that is, the system is open-loop unstable. Without any loss of generality, we assume A to be in Jordan form. Since we have assumed a Jordan form, we allow w_t to have correlated components, that is the correlation matrix $E[dd^T]$ is not necessarily diagonal. We also assume that B is invertible, even though, it suffices to only assume that (A, B) is a controllable pair.

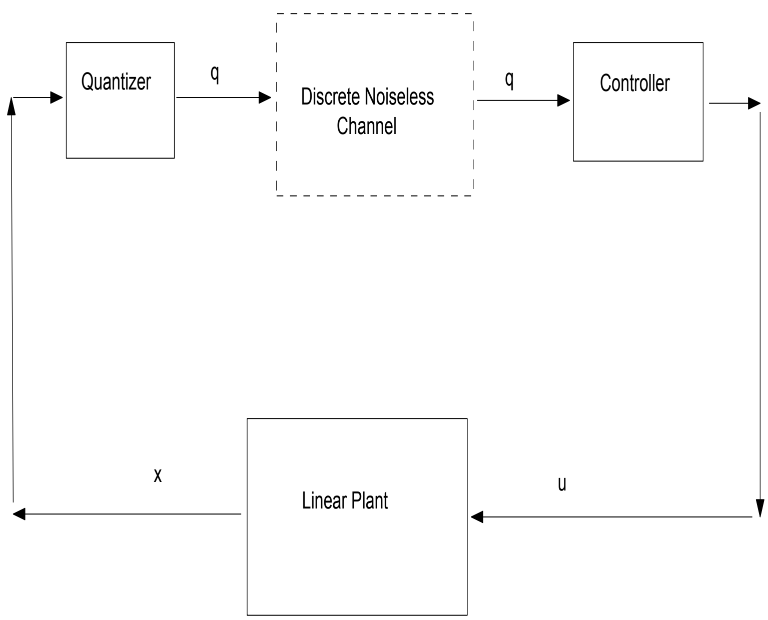

Figure 3: Control over a finite-rate noiseless channel with quantized observations at the controller.

The following is due to Wong-Brockett [161], Tatikonda-Mitter [141] and Nair-Evans [108].

Theorem 3.1 *[161] [141] [108] For any stabilizing control and causal quantization policy, if x_0 is second-order with finite entropy, for $\limsup_{t\to\infty} E[\|x_t\|^2] < \infty$, the average quantization rate needs to satisfy:*

$$R_{avg} \geq \sum_{|\lambda_i|>1} \log_2(|\lambda_i|),$$

where $\{\lambda_i, 1 \leq i \leq n\}$ denote the eigenvalues of A.

Proof: In the following proof, we use ideas from [177] and [181]. The matrix A can always be *block-diagonalized*, say with two blocks, where the first block has only stable eigenvalues, and the second one unstable eigenvalues. For the stable modes, one does not need to use the channel, and hence for the remaining discussion and analysis we can assume, without any loss of generality, that A has only unstable eigenvalues.

Let $x_t = E[x_t|q_{[0,t]}]$ and the Euclidean distance for the vector $x_t - \hat{x}_t$ be denoted by D_t, and the covariance matrix of the componentwise errors be denoted by C_t. Thus, $D_t = \text{Trace}(C_t)$. Let D_t be finite, which also makes C_t a matrix with finite entries. Among random vectors with a fixed covariance matrix, the differential entropy is maximized by a jointly Gaussian distribution, which in turn has a finite entropy [34]. It follows that

$$\begin{aligned}
\liminf_{T\to\infty}(1/T)H(q_{[0,T-1]}) &= \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1} H(q_t|q_{[0,t-1]}) + H(q_0)\Big) \\
&\geq \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(H(q_t|q_{[0,t-1]}) - H(q_t|x_t, q_{[0,t-1]})\Big) + H(q_0)\Big) \\
&= \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(I(x_t; q_t|q_{[0,t-1]})\Big) + H(q_0)\Big) \\
&= \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(h(x_t|q_{[0,t-1]}) - h(x_t|q_{[0,t]})\Big) + H(q_0)\Big) \\
&= \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1} + w_{t-1} + Bu_{t-1}|q_{[0,t-1]}) - h(x_t|q_{[0,t]})\Big) + H(q_0)\Big) \\
&= \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1} + w_{t-1}|q_{[0,t-1]}) - h(x_t|q_{[0,t]})\Big) + H(q_0)\Big) \\
&\geq \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1} + w_{t-1}|q_{[0,t-1]}, w_{t-1}) - h(x_t|q_{[0,t]})\Big) + H(q_0)\Big) \\
&= \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1}|q_{[0,t-1]}, w_{t-1}) - h(x_t|q_{[0,t]})\Big) + H(q_0)\Big) \\
&= \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(h(Ax_{t-1}|q_{[0,t-1]}) - h(x_t|q_{[0,t]})\Big) + H(q_0)\Big) \qquad (6) \\
&= \liminf_{T\to\infty}(1/T)\Big(\sum_{t=1}^{T-1}\Big(\log_2(|A|) + h(x_{t-1}|q_{[0,t-1]}) - h(x_t|q_{[0,t]})\Big) + H(q_0)\Big) \\
&= \liminf_{T\to\infty}(1/T)\Big(\Big(\sum_{t=1}^{T-1}\log_2(|A|)\Big) + h(x_0|q_0) - h(x_{T-1}|q_{[0,T-1]}) + H(q_0)\Big)\Big) \\
&\geq \liminf_{T\to\infty}(1/T)\Big(\Big(\sum_{t=1}^{T-1}\log_2(|A|)\Big) + h(x_0|q_0) - (1/2)\log((2\pi e)^n|C_{T-1}|) + H(q_0)\Big)
\end{aligned}$$

$$\geq \sum_{|\lambda_i|>1} \log_2(|\lambda_i|), \tag{7}$$

where C_t is the covariance matrix of the state error at time t. The first inequality follows since discrete entropy is always non-negative, and the second inequality follows from the fact that conditioning does not increase entropy. The third inequality follows from the property that the Gaussian measure maximizes the entropy, for a given covariance matrix. Equation (6) follows from the observation that $\{w_t\}$ is an independent process. The other equations follow from the properties of mutual information.

$\diamond$

3.1 Achievability through Random-time State-dependent Stochastic Drift

For achievability of the lower bound, we use a recently developed approach used first in [171], [168] and generalized in [180] [167]. Before, however, we provide a review of Markov chains and the notion of stochastic stability. See Meyn and Tweedie for a general discussion [100].

3.1.1 Stochastic Stability of Markov Chains

Let us first present a brief discussion on stochastic stability of Markov Chains; for a list of definitions on Markov Chains the reader is referred to [100] and [101]. Let $\{x_t, t \geq 0\}$ be a Markov chain with state space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, and defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{B}(\mathbb{X})$ denotes the Borel $\sigma-$field on $\mathbb{X}$, Ω is the sample space, $\mathcal{F}$ a sigma field of subsets of Ω, and $\mathcal{P}$ a probability measure. Let $P(x, D) := P(x_{t+1} \in D | x_t = x)$ denote the transition probability from x to D.

Definition 3.1 *For a Markov chain, a probability measure π is invariant on the Borel space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ if $\pi(D) = \int_{\mathbb{X}} P(x, D)\pi(dx), \quad \forall D \in \mathcal{B}(\mathbb{X})$.*

Definition 3.2 *A Markov chain is μ-irreducible, if for any set $B \in \mathcal{B}(\mathbb{X})$ with $\mu(B) > 0$, $\forall x \in \mathbb{X}$, there exists some integer $n > 0$, possibly depending on B and x, such that $P^n(x, B) > 0$, where $P^n(x, B)$ is the transition probability in n stages, that is $P(x_{t+n} \in B | x_t = x)$.*

Definition 3.3 *A set $A \subset \mathbb{X}$ is $\zeta-$petite on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ if for some distribution $\mathcal{Z}$ on $\mathbb{N}$ (set of natural numbers), and some non-trivial measure ζ, $\sum_{n=0}^{\infty} P^n(x, B)\mathcal{Z}(n) \geq \zeta(B), \quad \forall x \in A, \quad B \in \mathcal{B}(\mathbb{X})$.*

Definition 3.4 *A μ-irreducible Markov chain is aperiodic if for any $x \in \mathbb{X}$, and any $B \in \mathcal{B}(\mathbb{X})$ satisfying $\mu(B) > 0$, there exists $n_0 = n_0(x, B)$ such that $P^n(x, B) > 0$ for all $n \geq n_0$.*

Theorem 3.2 *[Meyn-Tweedie [101] Thm. 4.1] Suppose that $\boldsymbol{X}$ is a φ-irreducible Markov chain, and suppose that there is a set $A \in \mathcal{B}(\mathbb{X})$ satisfying the following:*

(i) *A is μ-petite for some μ.*

(ii) *A is recurrent: $P_x(\tau_A < \infty) = 1$ for any $x \in \mathbb{X}$.*

(iii) *A is* regular*: $\sup_{x \in A} E_x[\tau_A] < \infty$.*

Then $\boldsymbol{X}$ is positive Harris recurrent (and thus admits a unique invariant probability measure). $\square$

The existence of a unique invariant distribution is important also because of the following:

Theorem 3.3 (Birkhoff's Sample Path Ergodic Theorem) *Consider a positive Harris recurrent Markov process $\{x_t\}$ taking values in $\mathbb{X}$, with invariant distribution $\pi(.)$. Let $f : \mathbb{X} \to \mathbb{R}$ be such that $\int f(x)\pi(dx) < \infty$. Then, the following holds almost surely:*

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} f(x_t) = \int f(x)\pi(dx).$$

Throughout this section we consider a sequence of times $\{\mathcal{T}_i : i \in \mathbb{Z}_+\}$ which is assumed to be non-decreasing, with $\mathcal{T}_0 = 0$.

Theorem 3.4 *[Yüksel-Meyn [180][167]] Suppose that $\boldsymbol{X}$ is a μ-irreducible Markov chain. Suppose moreover that there is a function $V : \mathbb{X} \to (0, \infty)$, a petite set C, and constants $\kappa \in (0,1)$, $b \in \mathbb{R}$, such that the following hold:*

$$\begin{aligned} E[V(x_{\mathcal{T}_{z+1}}) \mid \mathcal{F}_{\mathcal{T}_z}] &\leq (1-\kappa)V(x_{\mathcal{T}_z}) + b1_{\{x_{\mathcal{T}_z} \in C\}} \\ E[\mathcal{T}_{z+1} - \mathcal{T}_z \mid \mathcal{F}_{\mathcal{T}_z}] &\leq V(x_{\mathcal{T}_z}), \qquad\qquad z \geq 0. \end{aligned} \tag{8}$$

Then, the Markov chain is positive Harris recurrent. □

The following provides a criterion for finite moments, which we refer to as random-time state-dependent stochastic drift.

Theorem 3.5 *[Yüksel-Meyn [180][167]] Suppose that $\boldsymbol{X}$ is a μ-irreducible Markov chain. Suppose moreover that there are functions $V : \mathbb{X} \to (0, \infty)$, $\delta : \mathbb{X} \to [1, \infty)$, $f : \mathbb{X} \to [1, \infty)$, a petite set C, and a constant $b \in \mathbb{R}$, such that the following hold:*

$$\begin{aligned} E[V(x_{\mathcal{T}_{z+1}}) \mid \mathcal{F}_{\mathcal{T}_z}] &\leq V(x_{\mathcal{T}_z}) - \delta(x_{\mathcal{T}_z}) + b1_{\{x_{\mathcal{T}_z} \in C\}} \\ E\Big[\sum_{k=\mathcal{T}_z}^{\mathcal{T}_{z+1}-1} f(x_k) \mid \mathcal{F}_{\mathcal{T}_z}\Big] &\leq \delta(x_{\mathcal{T}_z}), \qquad\qquad z \geq 0. \end{aligned} \tag{9}$$

Then $\boldsymbol{X}$ is positive Harris recurrent, and moreover $\lim_{t\to\infty} E[f(x_t)] = E_\pi[f(x)] < \infty$, with π being the invariant distribution. □

By taking $f(x) = 1$ for all $x \in \mathbb{X}$, we obtain the following corollary to Theorem 3.5.

Corollary 3.1 *Suppose that $\boldsymbol{X}$ is a φ-irreducible Markov chain. Suppose moreover that there is a function $V : \mathbb{X} \to (0, \infty)$, a petite set C, and a constant $b \in \mathbb{R}$, such that the following hold:*

$$\begin{aligned} E[V(x_{\mathcal{T}_{z+1}}) \mid \mathcal{F}_{\mathcal{T}_z}] &\leq V(x_{\mathcal{T}_z}) - 1 + b1_{\{x_{\mathcal{T}_z} \in C\}} \\ \sup_{x \in \mathbb{X},\ z \geq 0} E[\mathcal{T}_{z+1} - \mathcal{T}_z \mid x_{\mathcal{T}_z} = x] &< \infty. \end{aligned} \tag{10}$$

Then $\boldsymbol{X}$ is positive Harris recurrent. □

3.1.2 Scalar Case

Before proceeding further with the description of the system, we discuss the quantization policy investigated, for a scalar case. Consider the equation

$$x_{t+1} = ax_t + bu_t + w_t, \tag{11}$$

with $|a| > 1$ and $b \neq 0$, and w_t system disturbance.

We consider a variation of uniform quantizers. In the following, we modify the description of a traditional uniform quantizer by assigning the same value when the state is in the overflow region of the quantizer. As such, when $|x| > (K/2)\Delta$, the receiver knows that the source is in the overflow region of the quantizer. As such, a uniform quantizer: $Q_K^\Delta : \mathbb{R} \to \mathbb{R}$ with step size Δ and $K+1$ (with K even) number of bins satisfies the following for $k = 1, 2 \ldots, K$:

$$Q_K^\Delta(x) = \begin{cases} (k - \frac{1}{2}(K+1))\Delta, & \text{if } x \in [(k-1-\frac{1}{2}K)\Delta, (k - \frac{1}{2}K)\Delta) \\ (\frac{1}{2}(K-1))\Delta, & \text{if } x = \frac{1}{2}K\Delta \\ 0, & \text{if } x \notin [-\frac{1}{2}K\Delta, \frac{1}{2}K\Delta]. \end{cases}$$

A general class of quantizers are those which are adaptive. Let $\mathbb{S}$ be a set of states for a quantizer state S. Let $F : \mathbb{S} \times \mathbb{R} \to \mathbb{S}$ be a state update-function. An adaptive quantizer has the following state update equations: $S_{t+1} = F(Q_t(x_t), S_t)$. Here, Q_t is the quantizer applied at time t, x_t is the input to the quantizer Q_t, and S_t is the *state* of the quantizer. Such a quantizer is implementable since the updates can be performed at both the encoder and the decoder. One particular class of adaptive quantizers is introduced by Goodman and Gersho [51], which we will consider in the following analysis.

This system is connected over a noiseless channel with a finite capacity to an estimator (controller). The controller has access to the information it has received through the channel. The controller in our model estimates the state and then applies its control. As such, the problem reduces to a state estimation problem since such a scalar system is controllable. Hence, the stability of the estimation error is equivalent to the stability of the state itself.

An example of Goodman-Gersho [51] type adaptive quantizers, which also has been shown to be very useful in control systems, are those that have their bin sizes as the quantizer states [25]. In the zooming scheme, the quantizer enlarges the bin sizes in the quantizer until the state process is in the range of the quantizer, where the quantizer is in the *perfect-zoom* phase. Due to the effect of the system noise, occasionally the state will be in the overflow region of the quantizer, leading to an *under-zoom* phase. We refer to such quantizers as zooming quantizers. In the following, we will assume the communication channel to be a discrete noiseless one with capacity R.

Theorem 3.6 *[171] Consider an adaptive quantizer applied to the linear control system described by (19), under Assumption A. If the noiseless channel has capacity, for some $\epsilon > 0$, $R = \log_2(\lceil |a| + \epsilon \rceil + 1)$, there exists an adaptive quantization policy such that there exists a compact set S with $\sup_{x \in S} E[\min(t > 0 : x_t \in S)|x_0 = x] < \infty$, thus S is a regular [101] set.* ⋄

With K an even number, $R = \log_2(K+1)$, let us define $R' = \log_2(K)$. We will consider the following update rules. For $t \geq 0$ and with $\Delta_0 > L$ for some $L \in \mathbb{R}_+$, and $\hat{x}_0 \in \mathbb{R}$, consider:

$$u_t = -\frac{a}{b}\hat{x}_t, \quad \hat{x}_t = Q_K^{\Delta_t}(x_t), \quad \Delta_{t+1} = \Delta_t \bar{Q}\left(\left|\frac{x_t}{\Delta_t 2^{R'-1}}\right|, \Delta_t\right) \tag{12}$$

If we use $\delta > 0, 0 < \alpha < 1$ and $L > 0$ such that,

$$\begin{aligned} \bar{Q}(x, \Delta) &= |a| + \delta && \text{if} \quad |x| > 1 \\ \bar{Q}(x, \Delta) &= \alpha && \text{if} \quad 0 \leq |x| \leq 1, \Delta > L \\ \bar{Q}(x, \Delta) &= 1 && \text{if} \quad 0 \leq |x| \leq 1, \Delta \leq L, \end{aligned} \tag{13}$$

we will show that a recurrent set exists.

We now make explicit the connection with the general theory for random-time stochastic drift considered in Section 3.1.1. In the model considered, the controller can receive meaningful information regarding the state of the system when the source lies in the granular region of the quantizer: That is, $x_t \in [-\frac{1}{2}K\Delta_t, \frac{1}{2}K\Delta_t]$. The times at which these events occur form an increasing sequence of *stopping times*. We apply the drift criteria presented in Section 3.1.1 for these random stopping times. In particular, we will define $h_t := \frac{x_t}{\Delta_t 2^{R'-1}}$, observe that the process $\{(x_t, h_t)\}$ is Markov (equivalently $\{(x_t, \Delta_t)\}$ is Markov), and then define the sequence of stopping times as:

$$\mathcal{T}_0 = 0, \quad \mathcal{T}_{z+1} \;=\; \inf\{k > \mathcal{T}_z : |h_k| \leq 1\}, \quad z \in \mathbb{Z}_+. \tag{14}$$

These are the times when information reaches the controller regarding the value of the state when the state is in the granular region of the quantizer. We can establish criteria for recurrence for the Markov process $\{(x_t, h_t)\}$ (see Figure 4).

The following considers the state space for the quantizer bins to be countable, leading to irreducibility and consequently Positive Harris Recurrence.

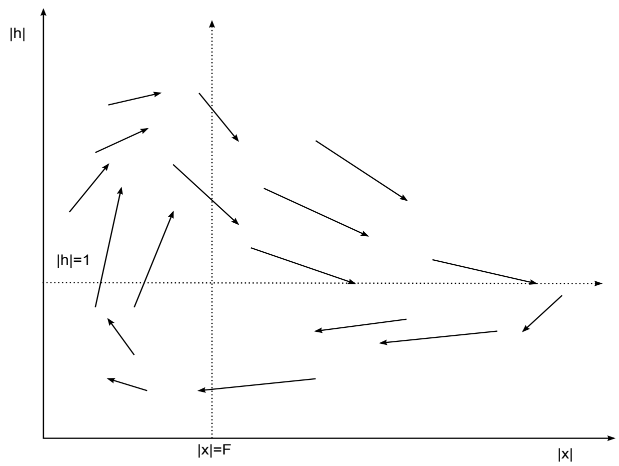

Figure 4: Drift in the Error Process. When under-zoomed, the error increases on average and the quantizer *zooms out*; when perfectly-zoomed, the error decreases and the quantizer *zooms in*.

Theorem 3.7 *Under the setup of Theorem 3.6, for the adaptive quantizer in (12), if the quantizer bin sizes are such that their (base–2) logarithms are integer multiples of some scalar s, and* $\log_2(\bar{Q}(\cdot,\cdot))$ *take values in integer multiples of s where the integers taken are relatively prime (that is they share no common divisors except for 1), then the process* $\{(x_t, \Delta_t)\}$ *is a positive (Harris) recurrent Markov chain, and, as such, has a unique invariant distribution.* ⋄

The following result is on moment stability.

Theorem 3.8 *[171] Under the setups of Theorem 3.6, Theorem 3.7 and Assumption B, it follows that* $\lim_{t\to\infty} E[x_t^2] < \infty$, *and this limit is independent of the initial state of the system.* ⋄

As a simulation study, we consider a linear system with the following dynamics:

$$x_{t+1} = 2.2x_t + u_t + w_t,$$

where $E[w_t] = 0, E[w_t^2] = 1$, and $\{w_t\}$ are i.i.d. Gaussian variables. We use the zooming quantizer with rate $\log_2(4) = 2$, since 4 is the smallest integer as large as $\lceil 2.2 \rceil + 1$. Figure 5 below corroborates the stochastic stability result, by showing the under-zoomed and perfectly-zoomed phases, with the peaks in the plots showing the under-zoom phases.

3.1.3 Multi-Dimensional Case

The proposed technique is also applicable for the multi-dimensional setup. Let us consider a multi-dimensional linear system

$$x_{t+1} = Ax_t + Bu_t + w_t, \tag{15}$$

where $x_t \in \mathbb{R}^n$ is the state at time t, u_t is the control input, and $\{w_t\}$ is a sequence of zero-mean independent, identically distributed (i.i.d.) $\mathbb{R}^n$-valued zero-mean Gaussian random variables. Furthermore, $E[||w_t||_2^2] = E[||w_1||_2^2] < \infty$. Here A is the system matrix with at least one eigenvalue greater than 1 in magnitude, that is, the system is open-loop unstable. Without any loss of generality, we assume A to be in Jordan form. Since we have assumed a Jordan form, we allow w_t to have correlated components, that is the correlation matrix $E[w_t w_t^T]$ is not necessarily diagonal. We also assume that B is invertible for making the stopping time analysis easier to pursue.

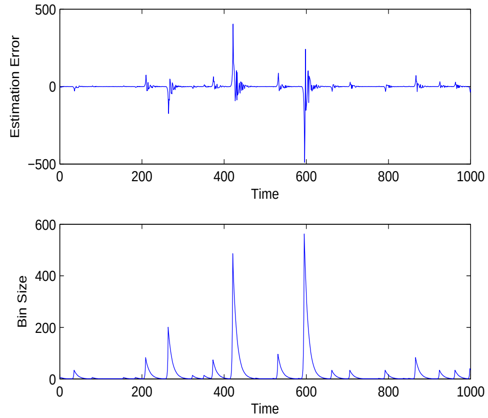

Figure 5: Sample path for a stochastically stable quantizer. The variables picked are as follows.

Instead of (14), the sequence of stopping times is defined as follows:

$$\mathcal{T}_0 = 0, \quad \mathcal{T}_{z+1} \;=\; \inf\{k > \mathcal{T}_z : |h^i_k| \leq 1, i = 1, 2 \ldots, n\}, \quad z \in \mathbb{Z}_+,$$

where $h^i_t = \frac{x^i_t}{\Delta^i_t 2^{R'_i - 1}}$. Here Δ^i is the bin size of the quantizer in the direction of the eigenvector x^i, with rate R'_i. With the above, the analysis follows that of scalar systems. For example, let us consider a two dimensional system with a Jordan form:

$$\begin{bmatrix} x^1_{t+1} \\ x^2_{t+1} \end{bmatrix} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} x^1_t \\ x^2_t \end{bmatrix} + B \begin{bmatrix} u^1_t \\ u^2_t \end{bmatrix} + \begin{bmatrix} w^1_t \\ w^2_t \end{bmatrix} \tag{16}$$

The approach is quantizing the components in the system according to the adaptive quantization rule provided earlier, that is, we modify the scheme in (12) as follows: Let for $i = 1, 2$, $R' = R'_i = \log_2(2^{R_i} - 1) = \log_2(K_i)$ (that is, the same rate is used for quantizing the components with the same eigenvalue). For $t \geq 0$ and with $\Delta^1_0, \Delta^2_0 \in \mathbb{R}$, consider:

$$u_t = -B^{-1} A \hat{x}_t,$$

$$\begin{bmatrix} \hat{x}^1_t \\ \hat{x}^2_t \end{bmatrix} = \begin{bmatrix} Q^{\Delta^1_t}_{K_1}(x^1_t) \\ Q^{\Delta^2_t}_{K_2}(x^2_t) \end{bmatrix}, \tag{17}$$

$$\Delta^i_{t+1} = \Delta^i_t \bar{Q}(|h^1_t|, |h^2_t|, \Delta^i_t), \quad i = 1, 2, \tag{18}$$

with $\delta^1 = \delta^2 > 0, 0 < \alpha^1 = \alpha^2 < 1$, and $L^i > 0 (i = 1, 2)$ such that

$$\begin{aligned} \bar{Q}^i(x, y, \Delta) &= |\lambda| + \delta^i && \text{if} \quad |x| > 1, \quad \text{or} \quad |y| > 1 \\ \bar{Q}^i(x, y, \Delta) &= \alpha^i && \text{if} \quad 0 \leq |x| \leq 1, |y| \leq 1, \Delta^i > L^i \\ \bar{Q}^i(x, y, \Delta) &= 1 && \text{if} \quad 0 \leq |x| \leq 1, |y| \leq 1, \Delta^i \leq L^i \end{aligned}$$

We also assume that for some sufficiently large η_Δ: $\Delta^1_0 = \eta_\Delta \Delta^2_0$, which leads to the result that $\Delta^1_t = \eta_\Delta \Delta^2_t$ for all $t \geq 0$.

With this approach, the drift criterion applies almost identically as it does for the scalar case.

We thus have the following result. See [74] for extensions of such a result.

Theorem 3.9 *Consider the multi-dimensional system (15). If the average rate satisfies*

$$R > \sum_{|\lambda_i|>1} \log_2(|\lambda_i|),$$

there exists a stabilizing scheme leading to a Markov chain with a bounded second moment in the sense that $\lim_{t\to\infty} E[||x_t||_2^2] < \infty$.

4 Quantizer Design for Optimization over Noiseless Channels: Separation Results

Consider the general setup described earlier in Section 3. Consider the setup earlier in (3) with one encoder.

Thus, the system considered is a discrete-time scalar system described by

$$x_{t+1} = f(x_t, w_t), \qquad y_t = g(x_t, r_t), \tag{19}$$

where x_t is the state at time t, and $\{w_t, r_t\}$ is a sequence of zero-mean, mutually independent, identically distributed (i.i.d.) random variables with finite second moments. Let the quantizer, as described earlier, map its information to a finite set $\mathcal{M}_t$. At any given time, the receiver generates a quantity v_t as a function of its received information, that is as a function of $\{q_0, q_1, \ldots, q_t\}$. The goal is to minimize $\sum_{t=0}^{T-1} E[c(x_t, v_t)]$, subject to constraints on the number of quantizer bins in $\mathcal{M}_t$, and the causality restriction in encoding and decoding.

4.0.4 Witsenhausen and Walrand-Varaiya's Separation Results

Let us revisit the single-encoder, fully observed case: In this setup, $y_t = x_t$ for all $t \geq 0$. There are two related approaches in the literature as presented explicitly by Teneketzis in [145]; one adopted by Witsenhausen [155] and one by Walrand and Varaiya [149]. Reference [145] extended the setups to the more general context of non-feedback communication.

Theorem 4.1 (Witsenhausen [155]) *An optimal causal composite quantization policy, if one exists, uses only* x_t *and* $q_{[0,t-1]}$ *at time* $t \geq 1$ *and* x_0 *at* $t = 0$.

Walrand and Varaiya considered sources living in a finite set, and obtained the following:

Theorem 4.2 (Walrand-Varaiya [149]) *An optimal causal composite quantization policy, if one exists, uses the conditional probability measure* $P(x_{t-1}|q_{[0,t-1]})$*, the state* x_t*, and the time information* t*, at time* $t \geq 1$.

The difference between the structural results above is the following: In Witsenhausen's setup, the encoder's memory space is not fixed and keeps expanding as the decision horizon in the optimization, $T-1$, increases. In Walrand and Varaiya's result, the memory space of an optimal encoder is fixed. In general, the space of probability measures is a very large one; however, it may be the case that different quantization outputs may lead to the same conditional probability measure on the state process, leading to a reduction in the required memory. Furthermore, Walrand and Varaiya's result allows one to apply the theory of Markov Decision Processes, for infinite horizon problems. We note that [19] applied such a machinery to obtain existence results for optimal causal coding of partially observed Markov processes.

4.1 Partially Observed Setting

Let for a general topological space $\mathbb{S}$, $\mathcal{P}(\mathbb{S})$ be the space of probability measures on $\mathcal{B}(\mathbb{S})$, the Borel $\sigma-$field on $\mathbb{S}$ (generated by opens sets in $\mathbb{S}$), equipped with the topology of total variation. Let $\pi_t \in \mathcal{P}(\mathbb{X})$ be the regular conditional probability measure (whose existence follows from the fact that both the state process and the observation process are complete, separable, metric, that is Polish, spaces) given by $P(dx_t|y_{[0,t]})$, that is

$$\pi_t(A) = P(x_t \in A|y_{[0,t]}), \quad A \in \mathcal{B}(\mathbb{X}).$$

It is known that the process $\{\pi_t\}$ evolves according to a non-linear filtering equation (see for example [63],[19]), and is itself a Markov process.

Let us also define $\Xi_t \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$ as the regular conditional measure

$$\Xi_t(A) = P(\pi_t \in A|q_{[0,t-1]}), \quad A \in \mathcal{B}(\mathcal{P}(\mathbb{X})).$$

The conditional measure $\{\Xi_t\}$ exists since the space of quantizer outputs is finite.

The following are our main results of this section:

Theorem 4.3 *[169][170] An optimal* causal composite quantization policy*, if there exists one, uses $\{\pi_t, q_{[0,t-1]}\}$ as a sufficient statistic for $t \geq 1$. This can be expressed as an optimal quantization policy which only uses $q_{[0,t-1]}$ to generate an optimal quantizer, where the quantizer uses π_t to generate the quantization output at time t.*

Theorem 4.4 *[169][170] An optimal* causal composite quantization policy*, if there exists one, uses $\{\Xi_t, \pi_t, t\}$ for $t \geq 1$. This can be expressed as an optimal quantization policy which only uses $\{\Xi_t, t\}$ to generate an optimal quantizer, where the quantizer uses π_t to generate the quantization output at time t.*

4.2 Application to Linear Gaussian Systems and Quadratic (LQG) Cost Minimization

Consider a Linear Quadratic Gaussian setup, where a sensor quantizes its noisy information to a controller. Let $x_t \in \mathbb{R}^n, y_t \in \mathbb{R}^m$, and the evolution of the system be given by the following:

$$\begin{aligned} x_{t+1} &= Ax_t + w_t, \\ y_t &= Cx_t + r_t. \end{aligned} \tag{20}$$

Here, $\{w_t, r_t\}$ is a mutually independent, white zero-mean Gaussian noise sequence with $W = E[w_t w_t^T], R = E[r_t r_t^T]$, A, B, C are matrices of appropriate dimensions. Suppose the the goal is the computation of

$$\inf_{\Pi^{composite}} \inf_{\gamma} E_{\nu_0}^{\Pi^{composite},\gamma}[\sum_{t=0}^{T-1} (x_t - v_t)^T Q(x_t - v_t)], \tag{21}$$

with ν_0 denoting a Gaussian distribution for the initial state, $Q > 0$ a positive definite matrix (See Figure 6).

The conditional distribution $\pi_t = P(dx_t|y_{[0,t]}, u_{[0,t-1]})$ is Gaussian for all time stages, which is characterized uniquely by its mean and covariance matrix for all time stages.

Theorem 4.5 *[169][170] For the minimization of the cost in (21) over all causal quantization policies, a causal quantizer based on the conditional distribution on the Kalman Filter output and the information available at the receiver is as good as any causal quantizer.*

We could, also obtain Walrand and Varaiya's structural result by considering a direct approach, exploiting the specific quadratic nature of the problem. Let, again, $x_t \in \mathbb{R}^n$ and $||.||$ denote the norm generated by an inner product of the form: $\langle x, y\rangle = x^T Q y$ for $x, y \in \mathbb{R}^n$ for positive-definite $Q > 0$. The Projection Theorem for Hilbert Spaces implies that the random variable $x_t - E[x_t|y_{[0,t]}]$ is orthogonal to the random variables $\{y_{[0,t]}, q_{[0,t]}\}$, where $q_{[0,t]}$ is included due to the Markov chain condition that $P(dx_t|y_{[0,t]}, q_{[0,t]}) = P(dx_t|y_{[0,t]})$. We thus obtain the following:

$$E[||x_t - E[x_t|q_{[0,t]}]||^2] = E[||x_t - E[x_t|y_{[0,t]}]||^2] + E[||E[x_t|y_{[0,t]}] - E[x_t|q_{[0,t]}]||^2]. \quad (22)$$

Thus, the optimality of Kalman filtering allows the encoder to only use the conditional estimate and the error covariance matrix without any loss of optimality (See Figure 6), and the optimal quantization problem also has an explicit formulation. Note that, the process $\{P(dx_t|y_{[0,t]})\}$ is Markov.

The above result is related to findings in [38] (also see [5] and [46]), and partially improves them in the direction of Markov sources.

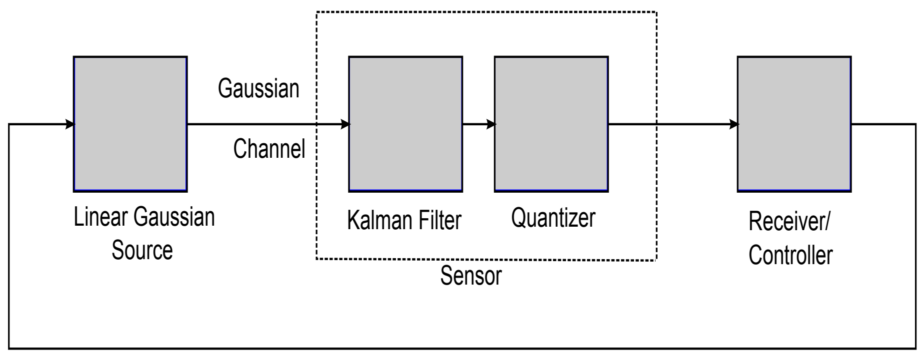

Figure 6: **Separation of Estimation and Quantization**: When the source is linear and Gaussian, the cost is quadratic, and the observation channel is Gaussian, the separated structure of the encoder above is optimal. That is, first the encoder runs a Kalman filter, and then causally encodes its estimate. For one-shot and independent observations setups, this result was observed in [5], [15], [12], [38], and [46]. Our result shows that, an extension of this result applies for the optimal causal encoding of partially observed Markov sources as well.

5 Quantizer Design for Decentralized Stabilization over Noiseless Channels

5.1 Stabilization of Linear Systems under Decentralized Information Structure

We now consider a class of multi-station n-dimensional discrete-time noise-free LTI systems

$$\begin{aligned} x_{t+1} &= Ax_t + \sum_{j=1}^{L} B^j u_t^j\,, \\ y_t^i &= C^i x_t\,,\;\; 1 \le i \le L, \end{aligned} \quad (23)$$

where $(A, [B^1|B^2|\dots|B^L])$ is controllable and $(A, [(C^1)^T|(C^2)^T|\dots|(C^L)^T]^T)$ is observable, but the individual pairs (A, B^i) may not be controllable or (A, C^i) may not be observable, for $1 \le i \le L$. Here, $x_t \in \mathbb{R}^n$, is the state of the system, $u_t^i \in \mathbb{R}^{m_i}$ is the control applied by station i, and $y_t^i \in \mathbb{R}^{p_i}$ is the observation available at station i at time t. Without any loss of generality, we assume the system matrix A to be in Jordan form.

The initial state x_0 is unknown, but is known to be generated according to some probability distribution which is supported on a compact set $\mathcal{X}_0 \subset \mathbb{R}^n$.

Under the decentralized information structure, the information available to station i at time t is

$$I^i_t = \{y^i_{[0,t]}, u^i_{[0,t-1]}\}$$

where, as before, $u^i_{[0,t-1]}$ denotes the sequence $\{u^i_0, u^i_1, \ldots, u^i_{t-1}\}$ and $y^i_{[0,t]} = \{y^i_0, y^i_1, \ldots, y^i_t\}$. We also assume that each of the stations has access to the system dynamics of all other agents.

Definition 5.1 *A decentralized system described in (23) is stabilizable under the decentralized information structure if there exists a set of admissible policies such that there exists a sequence* $\{K_t, t \geq 0\}$, $K_t \in \mathbb{R}$, $\lim_{t\to\infty} K_t = 0$ *with* $||x_t||_\infty \leq K_t \quad \forall t \geq 0$.

Let us define

$$\begin{aligned} \mathcal{K}^i &:= \begin{bmatrix} B^i & AB^i & \ldots & A^{n-1}B^i \end{bmatrix} \\ \mathcal{O}^i &:= \begin{bmatrix} (C^i)^T & (C^iA)^T & \ldots & (C^iA^{n-1})^T \end{bmatrix}^T \end{aligned}$$

and let the controllable and unobservable subspaces at station i be denoted by K^i and N^i, respectively, where K^i is the range space of $\mathcal{K}^i$ and N^i is the null-space of $\mathcal{O}^i$. We will, by a slight abuse of notation, refer to the subspace orthogonal to N^i as the observable subspace at the *ith* station and will denote it by O^i.

One of the important discussions in the literature is on the notion of *decentralized fixed modes*, for which there are two definitions. [36] presented a definition of fixed modes under linear time-invariant control policies: Let

$$\mathbb{F} = \{F : F = \mathrm{diag}(\mathrm{F}^1, \mathrm{F}^2, \ldots, \mathrm{F}^\mathrm{L}), \mathrm{F}^\mathrm{i} \in \mathbb{R}^{\mathrm{m_i}\times\mathrm{p_i}}\}.$$

The set of decentralized fixed modes under linear time-invariant laws is given by:

$$\Lambda = \bigcap_{\mathbb{F}} \lambda\left(A + \sum_{i=1}^{L} B^iF^iC^i\right),$$

where $\lambda(.)$ denotes the set of eigenvalues of its argument. Wang and Davison [36] proved that unless unstable fixed modes are present, a decentralized system can be stabilized by linear time invariant controllers. Anderson and Moore [2] provided algebraic conditions for there to be decentralized fixed modes under linear time-invariant policies; the reader is also referred to [105]. The other notion of fixed modes is on ones that are independent of the control policy applied, which arise due to the uncontrollability of the decentralized system.

Reference [79] showed that it is possible for the controllers to communicate through the plant, by signaling. We will provide further discussion on this in the development of the paper. If every station can communicate with every other station, possibly via other stations, the system is said to be strongly connected. Through communication via the plant, the controllable subspace can be expanded and the unobservable subspace can be reduced [79]. The work [2] showed that decentralized stabilization in a multi-controller setting is possible, if the system is jointly controllable, jointly observable, and strongly connected, via time-varying control laws. Corfmat and Morse [33] provided conditions for decentralized stabilization with time-invariant, output feedback controllers when a leader is picked to control the entire system. If a leader is selected, by restricting the other agents to use time-invariant or time-varying linear laws, the leader might be able to control the entire system under strong connectivity conditions.

Further related references are the works of Wang [150], Willems [152], Khargonekar and Özgüler [77] and Gong and Aldeen [50] which further studied time-varying control laws for stabilization. Khargonekar and Özgüler [77] studied the necessary and sufficient requirements needed for stabilization via time-varying controllers in terms of input output mappings. The conditions they provide is algebraic, and further corroborate the fact that strong connectivity does ensure decentralized stabilizability under the assumption of joint controllability and observability. Gong and Aldeen [50] considered the decentralized stabilization problem and obtained the characterization for stabilizability along similar algebraic conditions. Özgüner and Davison [118] used a sampling technique to eliminate fixed modes resulting from time-invariant policies.

The characterization of minimum information requirements for multi-sensor and multi-controller linear systems with an arbitrary topology of decentralization has been discussed in various publications and the discussion for the fundamental bounds has been extensively studied in [138], [139], [107] [109], [96], [97], [98], [99], [175], [176]. In particular, reference [97] introduced the idea of signaling to networked control problems, which was also studied in [176]. Recently, Matveev and Savkin provided a comprehensive discussion of the developments in networked control systems [99], and provided discussions on multi-controller systems when the eigenvalues are distinct, via cut-set type arguments. The case where either the plant or observation dynamics are noisy has been studied in [99] and [173]; where one common observation is that, information theoretic bounds for the noiseless case are not tight when the plant itself is noisy. Some other related results were presented earlier in [175] with time invariant policies and [138] for information theoretic considerations for multi-sensor systems. We finally refer the reader to earlier papers on control over communication channels for single-sensor and single-controller cases: [161], [143], [108] and [42].

The issue of complexity of decentralized computation is another important aspect for decentralized control applications. Wong [159] studied the communication complexity of decentralized control, building on the notion of *communication complexity of computation* by Yau [162]. In the information theory literature, distributed function computation with minimum information exchange is another important area, with some notable results being reported in [117]; which does not consider a real-time setup, but an information thoretic setup, which considers an infinite copy of messages to be encoded and functions to be computed, extending the results in Csiszar and Körner ([31], Thm. 4.6) to a computational setting. One could present two approaches for distributed computation for control systems: One is to allow the decision makers to share sufficient information to make the problem essentially centralized in a larger state space: This view had been adopted in [172], for the *belief sharing information pattern*. Another approach is to apply controls decentrally, this approach was adopted in [159].

Let $\mathbb{U} \subset \mathbb{R}^n$, $\mathbb{V} \subset \mathbb{R}^n$ be Euclidean subspaces. We adopt the following notation for such subspaces:

$$\begin{aligned} \mathbb{U} \cup \mathbb{V} &= \{v : v = \alpha a_1 + \beta a_2, a_1 \in \mathbb{U}, a_2 \in \mathbb{V}, \alpha, \beta \in \mathbb{R}\} \\ \mathbb{U} \cap \mathbb{V} &= \{v : v \in \mathbb{U}, v \in \mathbb{V}\} \\ \mathbb{U} - \mathbb{V} &= \{u : u \in \mathbb{U}, u^T v = 0, \forall v \in \mathbb{V}\} \end{aligned}$$

With the above definitions, for a vector space $S \subset \mathbb{R}^n$, we have $S^C = \mathbb{R} - S$, as the orthogonal complement of S. For vectors $v_1, v_2, \dots, v_m$, we denote by

$$\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_m}\} := \{\sum_{i=1}^{m} \alpha_i v_i, \alpha_i \in \mathbb{R}\},$$

the space spanned by the vectors $v_1, v_2, \dots, v_m$.

We denote by $\{v_1, v_2, \dots, v_m\}$ the set of individual vectors $v_1, v_2, \dots, v_m$.

For two sets Ψ and Ξ, $\Psi - \Xi = \{\eta : \eta \in \Psi, \eta \notin \Xi\}$, is the usual set difference.

We denote by $v_{[a,b]}$, for some vector v and $a, b \in \mathbb{Z}$, $a < b$, the sequence $(v_a, v_{a+1}, \dots, v_b)$.

By *modes* of a linear system, we refer to the subspaces (eigenspaces) which are invariant in the absence of control; as such, when all the eigenvalues of the system matrix are different the eigenvectors uniquely identify the modes of the system. In case the geometric multiplicity of an eigenvalue is less than its algebraic multiplicity, generalized eigenvectors (generalized modes) span the eigenspace corresponding to a particular eigenvalue.

We next introduce some relevant graph-theoretic notions: A *directed graph* $\mathcal{G}$ consists of a set of vertices, $\mathcal{V}$, and a set of directed edges, $(a, b) \in \mathcal{E}$, such that $a, b \in \mathcal{V}$. A path in $\mathcal{G}$ of length d consists of a sequence of d directed edges such that each edge is connected. A graph in which there exists a path from any node to any other is a strongly connected graph. We define the minimum distance between two sets of nodes

$S_1, S_2 \subset \mathcal{G}$ as $d(S_1, S_2) = \sum_{i \in S_1} \min\{d(i,j), j \in S_2\}$, where $d(i,j)$ denotes the number of paths between node i and j, with the trivial case being $d(i,i) = 0$ for all nodes.

5.1.1 Existence of Decentralized Controllers and Sufficiency of Restriction to Periodic Time-Varying Control Laws

In this section we discuss the necessary and sufficient conditions for there to be stabilizing controllers under the decentralized information structure considered in the paper.

Toward this end, we review the notion of connectivity. For two stations i and j, if $K^i \not\subseteq N^j$, then there exist control signals generated at station i which are observed at station j. We denote this by $i \to j$. This is equivalent to $C^j A^l B^i \neq 0$, for some $0 \leq l \leq n-1$ [176]. This ensures that station i can send information to station j through control actions $\{u_t^i\}$ (that is, engage in signaling). One may construct a directed communication graph given the above relation. It is possible that two stations are connected through other stations. If every station is connected to every other station, the system is said to be strongly connected. This, however, is a restrictive condition. Our result on the existence of stabilizing controllers is the following.

Theorem 5.1 *[176][99] Let the initial state satisfy $x_0 \in \mathcal{X}_0$, where $\mathcal{X}_0$ is a bounded set. The system (23) is controllable under the decentralized information structure IS if and only if there exist a partitioning of $\mathbb{R}^n$ in terms of $x^1, x^2, \ldots, x^n$ such that $\mathbb{R}^n = \cup_{i=1}^n \{\mathbf{x^i}\}$ and controllers such that for $i = 1, 2, \ldots, n$:*

$$x_0^i \in K^m \bigcap \left(\{ \bigcup_{l \to m} O^l \} \bigcup O^m \bigcup M_i \right)$$

where $M_1 = \{0\}$, and for $i > 1$,

$$M_i = \bigcup_{k=1}^{i-1} \{\mathbf{x_0^i}\}.$$

Kobayashi et al [79] presented a graph theoretic discussion where, they considered a case where the decentralized system can be expressed as a set of strongly connected sub-systems. They proved that the system is stabilizable by a linear controller if and only if there is no fixed mode between the decentralized system composed of the strongly connected subsystems. Khargonekar and Özgüler [77] made the connection with linear control laws. Let us revisit Theorem 2 of [77]. We first present a definition:

Definition 5.2 *Let a discrete time $n-$dimensional system with matrices A, B, C be given. This system is complete if*

$$\begin{bmatrix} \lambda I - A & B \\ -C & 0 \end{bmatrix}$$

has rank no smaller than n for all complex valued $\lambda \in \mathbb{C}$. If this holds for all $|\lambda| \geq 1$, then the system is said to be weakly complete.

In Lemma 1 of [77], it is proven that if a system is complete, and if $CA^lB = 0$ for all $l \geq 0$, then the controllable subspace of (A, B) is identical to the unobservable subspace of (A, C).

Theorem 5.2 *[77] There exists a periodic-time-varying decentralized controller if and only if the joint system is stabilizable and detectable, and for every partitioning of the system into $\mathcal{E} = a_1, a_2, \ldots, a_k$, $\mathcal{F} = b_1, b_2, \ldots, b_{L-k} =: \{1, 2, \ldots, L\} - \mathcal{E}$, such that if the system*

$$(A, [B^{a_1} B^{a_2} \ldots B^{a_k}], [(C^{b_1})^T (C^{b_2})^T \ldots (C^{b_{L-k}})^T]^T),$$

has a zero transfer function, then it is weakly complete.

We now state a theorem on the universality of linear-time varying controllers for decentralized stabilization.

Theorem 5.3 *[50] A decentralized system described in (23) is stabilizable under the decentralized information structure if and only if it is stabilizable by periodic linear time-varying controllers.*

5.2 Decentralized Stabilization under Quantization Constraints and Signaling

Suppose the controllers are connected to the plant over a discrete-noiseless channel. In this case, the control signals u^i are coded and decoded over discrete noiseless channels with finite capacity. Hence, the applied control and transmitted messages follow a coding (i.e., binary representation) and a decoding process. We assume fixed-rate encoding, that is, the rate is defined as the (base-2) logarithm of the number of symbols to be transmitted: The coding process of the controller at station i is a mapping measurable with respect to the sigma-algebra generated by I_t^i to $\{1, 2, \ldots, W_t^i\}$, which is the quantizer codebook at station i at time t. Hence, at each time t, station i sends $\log_2(W_t^i)$ bits over the channel to the plant.

Problem Statement Let $\mathcal{R}$ denote the set of average rates on L sensor and controller channels which lead to stabilization, that is,

$$\mathcal{R} = \Big\{ R^i, i \in 1, 2, \ldots, L : \\ \exists \{u^1_{[0,\infty)}, u^2_{[0,\infty)}, \ldots, u^L_{[0,\infty)}\}, \lim_{T\to\infty} ||x_T||_\infty = 0 \Big\}$$

where $R^i = \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \log_2(W_t^i)$. We are interested in the *average total rate* $\mathbf{R} := \min_{\mathcal{R}} \{\sum_{i=1}^{L} R^i\}$, such that decentralized stabilization is possible. ⋄

Two supportive results are in order. The first one will provide the rate needed for the controllers and sensors to communicate the necessary information to the controllers capable of controlling a mode. The second result will investigate the rate from the controllers to the plant itself.

5.2.1 Case where Multiple Controllers Need information from Different Controllers/Sensors

Let there be more than one controllers which can control a mode, yet, their information is not sufficient to recover the mode independently. We remark that this may apply only for some modes with geometric multiplicities greater than one. One example is the following:

$$x_{t+1} = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u_t^1 + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u_t^2 + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} u_t^3 + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} u_t^4$$

$$y_t^1 = [0\ 0\ 0\ 1]\, x_t$$
$$y_t^2 = [0\ 0\ 1\ 0]\, x_t$$
$$y_t^3 = [1\ 1\ 0\ 0]\, x_t$$
$$y_t^4 = [1\ \ -1\ 0\ 0]\, x_t$$

In this case, the third and fourth stations send information to the first two controllers, which can control the first mode whose information is not enough to recover the mode independently.

Theorem 5.4 *The minimum average total information rate needed to be sent to the controllers for controlling a mode with eigenvalue λ_i is lower bounded by*

$$\min_{\mathbb{K},\mathbb{L}:\{\mathbf{x^i}\}\subset(\cup_{m\in\mathbb{K}} K^m)\cap(\cup_{j\in\mathbb{L},m\in\mathbb{K}} O^j \cup O^m)} d(\mathbb{L},\mathbb{K}) \max(0, \log_2(|\lambda_i|)), \tag{24}$$

where

$$d(\mathbb{L},\mathbb{K}) = \sum_{l\in\mathbb{L}} \min_{k\in\mathbb{K}} d(l,k).$$

5.2.2 Case where Multiple Controllers Jointly Control a Given Mode

We now discuss how much information is needed to be transmitted from the controllers to the plant to apply the control action. Let there be a number of controllers who can control a given mode, but only their joint information is sufficient to recover the mode. In this case:

Theorem 5.5 *The minimum average total information rate needed to be sent from the controllers to the plant for controlling a mode with eigenvalue λ_i is lower bounded by*

$$\min_{\mathbb{K}:\{\mathbf{x^i}\}\subset(\cup_{m\in\mathbb{K}}K^m)\cap(\cup_{m\in\mathbb{K}}O^m)} \max(0,\log_2(|\lambda_i|))|\mathbb{K}|$$

Remark: If only one controller applies the control, then the price of communication is lowest. But it might be more beneficial to communicate to two separate different controllers. When the eigenvalues are all distinct, then it suffices to only consider one controller, as there will be a controller to whom the information can be sent most efficiently. The issue of different controllers acting will arise when the algebraic multiplicity of a mode is greater than one. ◇

5.2.3 Tight Lower Bound on the Minimum Average Total-Rate

The following result follows from [176]. In the following, we restrict the analysis to the case when the eigenvalues are distinct, although a similar analysis is possible for the case when the eigenvalues are repeated, requiring further definitions.

Theorem 5.6 *[176] Let A be such that all the eigenvalues are distinct and let $\{x^i\}$ be the corresponding eigenvectors. Then, a tight lower bound on the total rate required,* $\mathbf{R}$*, between the controllers and the plant for stabilizability is given by*

$$\Bigg\{\sum_{|\lambda_i|>1}(\eta_{M_i})\bigg(\log_2(|\lambda_i|)\bigg)\Bigg\}, \tag{25}$$

where

$$\eta_{M_i} = \min_{l,m\in\{1,2,\dots,L\}}\Bigg\{D^*(l,m): l\to m, \{\mathbf{x^i}\}\subset O^l, \\ \{\mathbf{x^i}\}\subset K^m, D^*(l,m)=d(l,m)+1.\Bigg\} \tag{26}$$

Proof: If $k\to l$, then station k can send the information on a particular mode in O^k to station l, which, upon receiving the state information can generate the control signal. For this we use quantization, where successively the initial state is sent to the controller. Without any loss of generality, suppose that there exists only one station, station l, that can control a mode i, and only one station, $k\neq l$, that can observe a mode x^i. Then, the information on mode x^i is to be sent to station l through the plant; thus the plant *relays* the information.

Suppose x_0^i is to be sent to station l. Sensor k recovers x_0^i at a time no later than n. It then quantizes x_0^i uniformly. Station k sets $u_t^k = Q_t(x_0^i)$, where Q_t denotes the quantizer used at time t. In this case,

$$x_{n+1} = Ax_n + B^kQ_n(x_0^i),$$
$$y_{n+1}^l = C^l(Ax_{n+1} + B^kQ_n(x_0^i)).$$

Assembling the observations $y_{n+1}^l, y_{n+2}^l, \dots, y_{2n}^l$, and following the fact that $C^l(A)^mB^k\neq 0$ for at least one $m, 1\leq m\leq n$, the quantized output $Q_n(x_0^i)$ can be recovered at a time no later than $2n$.

Thus, sensor l can recover the quantized information $Q_n(x_0^i)$, which it subsequently sends to station l. Via this information, the estimate at time $2n$, $\hat{x^i}_0(2n)$, can be computed.

Let $p > 0$ be an integer. If an average quantization rate of $R = n \log_2 |\lambda_i| + \epsilon$, for some $\epsilon > 0$ is used, then the estimation error $x_0^i - \hat{x^i}_0(pn)$ approaches zero at a rate faster than

$$|x_0^i - \hat{x}_0^i(pn)| < 1/(|\lambda_i|)^{pn}.$$

The plant undoes the signaling, since it is assumed to know the control protocol. Since the system is controllable, the controller can drive the estimated value to zero in at most n time stages. The rate required for the transmission of control signals will be identical to the rate of the transmission of the sensory information, since the control operation will be a one-to-one mapping, which is in fact linear in the observations.

Finally, we need to consider multiple transmissions. The remaining controllers can be designed to be idle, while a particular mode is being relayed by the plant. Such a sequential scheme ensures convergence. Since the number of time-stages are adjustable, ϵ can be taken to converge to zero. ◇

5.3 Incorporation of Sensors into the Model

We can also assume there to be sensors in the system. Suppose there are a number of sensors which are directly connected to a number of controllers. We can efficiently model such a sensor as a control station which can only control one artificial, decoupled, stable mode, and use this stable mode to communicate signals to the stations that the sensors can communicate to. In this case, the enlarged state will have a larger dimension, but will not affect the unstable subspace of the system as the new components will be decoupled. It is further possible that these sensors are signaled by other controllers. An example is the following. Consider a three controller system, where there is also an additional sensor which is connected to station 1,

$$x_{t+1} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -2 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_t^1 + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_t^2 + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} u_t^3$$

$$y_t^1 = [0\ 0\ 0]\, x_t$$
$$y_t^2 = [0\ 0\ 1]\, x_t$$
$$y_t^3 = [1\ 1\ 0]\, x_t$$

with the sensor observation as:

$$y_t^S = [0\ 0\ 1]\, x_t$$

We can express this system as a multi-controller system by regarding the sensor as a new station, station 4, which cannot stabilize an unstable mode but can signal information through a fictitious stable mode it can control. The new four-controller system can be constructed as follows, where the new mode only acts as a communication channel between station 4 and station 1:

$$x_{t+1} = \begin{bmatrix} 5 & 1 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u_t^1 + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u_t^2 + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} u_t^3 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} u_t^4$$

$$y_t^1 = [0\ 0\ 0\ 1]\, x_t$$
$$y_t^2 = [0\ 0\ 1\ 0]\, x_t$$
$$y_t^3 = [1\ 1\ 0\ 0]\, x_t$$
$$y_t^4 = [0\ 0\ 1\ 0]\, x_t$$

The above construction, together with Theorem 5.2 leads to the following result.

Proposition 5.1 *In system (23) with additional sensors, if the sensors are incorporated in the system as controllers which can control a fictious stable mode used for signaling to a corresponding controller, then the system will be weakly complete if the original system is weakly complete.*

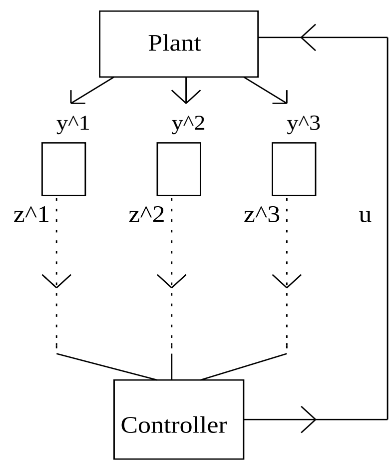

Figure 7: Multisensor system structure.

5.4 Multi-Sensor Structure with a Centralized Controller as a Special Case

It should be noted that, the general problem considered in this paper is not a multi-terminal source coding problem with a centralized decoder where a number of remote sensors encode information to a centralized decision maker (fusion center). This is why there is a rate loss when compared with a system with a single controller and a single sensor. The decoder (some controller) might have the observation information to control all the modes as a result of signaling, but cannot control all such modes, but requires to receive all of them to extract what she needs. Hence, it is unavoidable not to send this information to the controller decentrally. In a multi-terminal source coding problem, there is no-rate loss, because, the additional information received is also related and can be utilized. This is in contrast with the multi-sensor problem, with a centralized controller, for which it is known that there is no rate loss (see [96], [138], [174], [109]), which is in fact a special case of the multi-controller problem as discussed earlier. We now revisit this problem. Consider the following discrete-time system (see Figure 7)

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t\,,\ t \geq 0, \\ y_t^i &= C^i x_t \end{aligned} \tag{27}$$

where (A, B) is stabilizable, x_t is the state, u_t is the control, and the initial state x_0 is a random vector with a known continuous distribution over a compact support. Assume $(A, [(C^1)^T \ldots (C^L)^T]^T)$ to be detectable. The information sent by the sensors is quantized and is sent to the controller, which has access to only the information sent by the sensors.

The geometric conditions presented in the paper shows that the rate required is the sum of the logarithms of the unstable eigenvalue magnitudes, see [138], [174], [98] for related results.

6 Quantizer Design for Optimal Decentralized Control over Noiseless Channels

We now consider the optimization of quantizers in a decentralized system. In view of the non-classical information structure, it is not always possible to obtain optimality results for a general setup; however, for a class of information structures, it is possible to obtain a characterization of optimal quantizers.

One question is the following: Suppose that there is a channel between two controllers. What is the performance of the decentralized system, given a rate constraint on the communication channel. That is, consider the following problem:

Under an ensemble of quantization and control policies $\mathbf{\Pi}$ and a given information pattern, with an initial condition x_0, the attained performance index is

$$J_{x_0}(\mathbf{\Pi}) = E_{x_0}^{\mathbf{\Pi}}[\sum_{t=0}^{T-1} c(x_t, \mathbf{u}_t)]$$

subject to constraints

$$\sum_{t=0}^{T-1} R_t^{i,j} \leq \bar{R},$$

where $R_t^{i,j}$ denotes the bit rate per time stage for communication among the decision makers (see Fig. 8).

For the case when $\bar{R}$ allows for exchange of *agreement on the conditional measure on the state*, we can obtain a precise answer, which we do in the following.

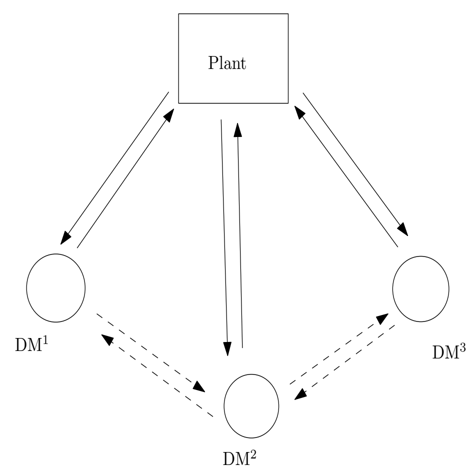

Figure 8: Agreement on the conditional measure on the state

In the following, let the state and observation spaces be finite.

6.1 One-Stage Belief Sharing Pattern

In the following, we use ideas from [163] and [172]. Let there be a common information vector I_t^c at some time t, which is available at all of the decision makers. Let at times $ks-1$, $k \in \mathbb{Z}_+ \cup \{0\}$ and T divisible by k, $s \in \mathbb{Z}_+$, the decision makers share all their information: $I_{ks-1}^c = \{\mathbf{y}_{[0,ks-1]}, \mathbf{u}_{[0,ks-1]}\}$ and for $I_0^c = \{P(x_0)\}$, that is at time 0 the DMs have the same apriori belief on the initial state. Until the next observation instant $t = k(s+1)-1$ we can regard the individual decision functions specific to DM^i as $\{u_t^i = \bar{u}_s^i(y_{[ks,t]}^i, I_{ks-1}^c)\}$ and we let $\bar{\mathbf{u}}$ denote the ensemble of such decision functions. In essence, it suffices to generate $\bar{\mathbf{u}}_s$ for all $s \geq 0$, as the decision outputs conditioned on $y_{[ks+1,t]}^i$, under $\bar{u}_s^i(y_{[ks,t]}^i, I_{ks-1}^c)$, can be generated.

In such a case, we have that $\bar{\mathbf{u}}_s(., I_{ks-1}^c)$ is the joint team decision rule mapping I_{ks-1}^c into a space of action vectors: $\{u^i(I_{ks-1}^c, y_{[ks,t]}^i), i \in \{1, 2 \ldots, L\}, t \in \{ks, ks+1, \ldots, k(s+1)-1\}\}$. In this case, the cost function is also modified as:

$$J_{x_0}(\mathbf{\Pi}) = E_{x_0}^{\mathbf{\Pi}}[\sum_{s=0}^{\frac{T}{k}-1} \bar{c}(\bar{\mathbf{u}}_s(., I_{ks-1}^c), \bar{x}_s)]$$

with

$$\bar{c}(\bar{\mathbf{u}}_s(., I_{ks-1}^c), \bar{x}_s) = E_{\bar{x}_s}^{\mathbf{\Pi}}[\sum_{t=ks}^{k(s+1)-1} c(x_t, \mathbf{u}_t)]$$

Lemma 6.1 *[172] Consider the decentralized system setup in Section 1, with the observation noise processes being independent. Let I^c_t be a common information vector supplied to the DMs regularly at every k time stages, so that the DMs have common memory with a control policy generated as described above. Then, $\{\bar{x}_s := x_{ks}, \bar{\mathbf{u}}_s(., I^c_{ks-1}), s \geq 0\}$ forms a Controlled Markov chain*

In view of the above, we now present a result on a separation property. We note that the following also has been studied in [113]. We present a shorter proof, using the result above directly.

Lemma 6.2 *[172][113] Let I^c_t be a common information vector supplied to the DMs regularly at every k time steps. There is no loss in performance if I^c_{ks-1} is replaced by $P(\bar{x}_s|I^c_{ks-1})$.*

The essential issue for a tractable solution is to ensure a common information vector which will act as a sufficient statistic for future control policies. This can be done via sharing information at every stage, or some structure possibly requiring larger but finite delay.

Definition 6.1 Belief Sharing Information Pattern: *[172] An information pattern in which the DMs share their beliefs about the system state is called the* belief sharing information pattern. *If the belief sharing occurs periodically at every k-stages ($k > 1$), the DMs also share the control actions they applied in the last $k-1$ stages, together with intermediate belief information. In this case, the information pattern is called the* k-stage belief sharing information pattern. ⋄

Let us consider the one-stage belief sharing pattern, first for a two DM setup. In this case, the information needed at both the controllers is such that they all need to exchange the relevant information on the state, and need to agree on $p(\bar{x}_t|I^1_t, I^2_t)$, where I^i_t denotes the information available at DM^i. In the one-step Belief Sharing Pattern, $\bar{x}_t = x_t$, since the period for information exchange $k = 1$.

We note that, when control policies are deterministic, the actions can uniquely be identified by both DMs. As such, control signals need not be exchanged.

Theorem 6.1 *[172] To achieve the one-stage belief sharing information pattern, the following rate region is achievable using fixed-rate codes:*

$$\begin{aligned}\mathcal{R}(t) \quad = \quad & \Bigg\{(R^{i,j}, R^{j,i}) : R^{i,j} = \lceil \log_2(|\mathcal{S}_t|) \rceil, R^{j,i} = \lceil \log_2(\sup_{\pi^i} |\mathcal{S}_{\pi^i,t}|) \rceil, \\ & \quad \mathcal{S}_t = \left\{ \pi^i = P\left(x_t \middle| y^i_t = y^1, P(.|.)\right) : P\left(y^i_t = y^i \middle| P(.|.)\right) > 0, y^i \in \mathbb{Y}^i \right\}, \\ & \quad \mathcal{S}_{\pi^i,t} = \left\{ P\left(x_t \middle| y^j_t = y^j, \pi^i, P(.|.)\right) : P\left(y^j_t = y^j \middle| \pi^i, P(.|.)\right) > 0, y^j \in \mathbb{Y}^j \right\} \Bigg\},\end{aligned}$$

where $P(.|.)$ denotes $P(x_t|I^c_{t-1})$.

A discussion is also available when the communicate rate is measured by the average number of bits.

Theorem 6.2 *Suppose the observation variables are discrete valued, that is $\mathbb{Y}^i, i = 1, 2$ is a countable space. To achieve the belief sharing information pattern, a lower bound on the minimum average amount of bits to be transmitted to DM^i from DM^j, $i, j \in \{1, 2\}, i \neq j$ and in the opposite direction are*

$$R^{j,i} \geq H\left(P(x_t|I^c_{t-1}, y^i_t, y^j_t) \middle| P(x_t|I^c_{t-1}), y^i_t \right),$$

$$R^{i,j} \geq H\left(P(x_t|I^c_{t-1}, y^i_t, y^j_t) \middle| P(x_t|I^c_{t-1}), y^j_t, Z^i_t \right),$$

where Z^i_t is the variable sent to DM^1 from DM^2.

We note that the information rate needed is less than one needed for achieving the centralized information pattern. By the above argument, one would need $R^{i,j} \geq H(y_t^i | y_t^j, I_{t-1}^c)$ for the centralized information pattern as a lower bound. The entropy of the conditional measure is at most as much as the entropy of the observed variable. This is because, different outputs may lead to the same values for $P(y_t^2 = y | x_t, I_{t-1}^c)$. Hence, we have the following corollary to Theorem 6.2.

Corollary 6.1 *When the observation space is discrete, the one-stage belief sharing information pattern requires less or equal amount of information exchange between the controllers than the centralized information pattern.*

We may also obtain an upper bound on the communication rates for a two decision maker setting for variable-rate schemes.

Proposition 6.1 *To achieve the one-stage belief sharing information pattern, an upper bound on the minimum average amount of bits to be transmitted to DM^i from DM^j, $i, j \in \{1, 2\}, i \neq j$, is given by:*

$$R^{j,i} \leq \min \Bigg\{ H\bigg(\zeta(y_t^j, P(.|.)) \bigg| P(.|.) \bigg) : \\ P\bigg(x_t \bigg| P(.|.), y_t^1, y_t^2 \bigg) = P\bigg(x_t \bigg| P(.|.), y_t^i, \zeta(y_t^j, P(.|.)) \bigg) \Bigg\},$$

where $P(.|.)$ denotes $P(x_t | I_{t-1}^c)$.

Remark: In the setup considered, the goal is that each DM can compute the joint belief. We note here the interesting discussion between decentralized computation and communication provided in Csiszar and Körner ([31], Thm. 4.6) and Orlitsky and Roche [117]. However, the setting presented in these works assumes an infinite copy of messages to be encoded and functions to be computed, which is not applicable in a real-time setting. ◇

7 Concluding Remarks and Some Open Problems

This tutorial left out many problems and questions; in view of conciseness and space constraints.

7.1 Quantizer Design for Optimal Control over Erasure and Noisy Channels

The erasure channel is an important practical channel, and one could argue that it is the most relevant noisy channel for real-time control applications. For such channels, both stabilization and optimization have been considered in the literature. We refer the reader to [71], [103], [164], [180], [167], [132], [60], [166].

Likewise, for discrete-alphabet as well as continuous-alphabet noisy channels, there has also been a significant activity. References [99], [143], [181], [145], [87] and [7] contain a rich literature review. There still remain many open problems, in particular on synchronization across noisy channels, and optimal encoder design with noisy feedback. The book [99] is an excellent reference for a comprehensive discussion of general channels and a literature review.

7.2 Existence of Optimal Quantizers and Optimal Quantization of Probability Measures

The problem of existence of optimal quantizer, and topological properties of information channels in networked control problems has been addressed in a number of papers, such as [19] and [179].

In [179] existence of optimal quantizers is addressed for the case when the source admits a probability density. One further problem is on the existence and design of optimal quantizers. There are results available in the literature on optimal quantization of probability measures [53], however, more needs to be done in the context of finite rate communications and for the cases when the source lives in an uncountable space. With a separation result paving the way for an MDP formulation, one could proceed with the analysis of [19] with the evaluation of the optimal quantization policies and existence results for infinite horizon problems. Nonetheless, quantization of probability measures remains an interesting problem to be understood in a real-time coding context; with important practical consequences. In [179], existence of optimal quantizers has been studied for finite dimensional Euclidean state spaces, when a source distribution admits a density.

The separation results in the paper and in the literature will likely find many applications in sensor networks and networked control problems where sensors have imperfect observation of a plant to be controlled. One direction is to find explicit results on the optimal policies using computational tools. One promising approach is the expert-based systems, which are very effective once one imposes a structure on the designs, see [62] for details.

7.3 Optimal Information Exchange and Information Structures with an Information Theoretic Characterization

One related problem is how to optimally encode the control actions; such an analysis will explicitly depend on the cost functions [146]. In this case, the exchanged information exhibits the properties of the triple effect of control; as such the analysis requires further research.

In the paper, we discussed the case when the rate is high enough to allow belief sharing. The regime where this is not possible requires further research, since the exchange information affects both the current stage cost and the costs in future stages. Deterministic nestedness has been observed to be too restrictive in [172]. One observes that it is possible to transform an information structure into a desirable one leading to tractable decentralized optimization problems by exchange of minimal information. Such an analysis will have important practical consequences. The dynamic nature of such quantization problems, however, require first a precise notion of *decentralized state of a system.*

8 Acknowledgements

The author gratefully acknowledges collaborators and co-authors Tamer Başar, Sean P. Meyn, Sekhar Tatikonda, Orhan C. Imer, Tamás Linder, Aditya Mahajan, Giacomo Como, Maxim Raginsky and Andrew Johnston for extensive discussions and their collaborations on much of the contents presented in this article.

References

[1] M. Aicardi, F. Davoli, R. Minciardi, Decentralized optimal control of Markov chains with a common past information set, IEEE Trans. Automatic Cotrol, V. 32, N. 11, pp. 1028-1031, 1987.

[2] B. D. O. Anderson and J. B. Moore, Time-varying feedback decentralized control, IEEE Trans. Automatic Control, V. 26, N. 10, pp. 1133-1139, 1981.

[3] A. Arapostathis, V. S. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh and S. I. Marcus, Discrete-Time Controlled Markov Processes with Average Cost Criterion: A Survey, SIAM J. Control and Optimization, V. 31, pp. 282-344, 1993.

[4] M. Aoki, On decentralized linear stochastic control problems with quadratic cost, IEEE Trans. Automatic Control, V. 18, N. 6, pp. 243-250, 1973.

[5] E. Ayanoğlu and R. M. Gray, The design of joint source and channel trellis waveform coders, IEEE Trans. on Inform. Theory, V. 33, N. 6, pp. 855865, Nov. 1987.

[6] B. Bamieh and P. Voulgaris, A Convex Characterization of Distributed Control Problems in Spatially Invariant Systems with Communication Constraints, Systems and Control Letters, V. 54, pp. 575-583, 2005.

[7] R. Bansal and T. Başar, Simultaneous design of measurement and control strategies for stochastic systems with feedback, Automatica, V. 25, pp. 679-694, 1989.

[8] Y. Bar-Shalom, and E. Tse, Dual effect certainty equivalence and separation in stochastic control, IEEE Trans. Automatic Control, V. 19, N. 10, p. 494500, 1974.

[9] T. Başar, Decentralized multicriteria optimization of linear stochastic systems, IEEE Trans. on Automatic Control, V. 23, N. 4, pp. 233 - 243, 1978.

[10] T. Başar and J. Cruz, Concepts and methods in multiperson coordination and control, in Optimization and Control of Dynamic Operational Research Models (Editor: S. G. Tzafestas), Chapter 11, pp. 351 - 387, North Holland, 1982.

[11] T. Başar, Control and game-theoretic tools for communication networks (Overview), Applied and Computational Math., V. 6, N. 2, pp. 104-125, 2007.

[12] R. Bansal and T. Başar, Solutions to a class of linear-quadratic-Gaussian LQG stochastic team problems with nonclassical information, Systems and Control Letters, V. 9, pp. 125 - 130, 1987

[13] J. M. Bismut, An example of interaction between information and control: The transparency of a game, presented at the Joint Harvard University - Imperial College Conf. Information Structures and Extensive Games, June 1972.

[14] D. Blackwell, Equivalent comparison of experiments, Annals of Mathematical Statistics, V. 24, pp. 265272, 1953.

[15] T. Berger, Rate-Distortion Theory, Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[16] T. Berger, Information rates of Wiener processes, IEEE Trans. Inform. Theory, V. 16, pp. 134-139, March 1970.

[17] D. P. Bertsekas, Dynamic Programming and Stochastic Optimal Control, Academic Press, New York, New York, 1976.

[18] R. E. Blahut, Principles of Information Theory, Addison-Wesley, Reading, MA, 1987.

[19] V. S. Borkar, S. K. Mitter, and S. Tatikonda, Optimal sequential vector quantization of Markov sources, SIAM J. Control and Optimization, V. 40, pp. 135-148, 2001.

[20] V. S. Borkar, Probability Theory: An Advanced Course, Springer, New York, 1995.

[21] V. Borkar and P. Varaiya, Asymptotic agreement in distributed estimation, IEEE Trans. Automatic Cotrol, V. 27, pp. 650 - 655, June 1982.

[22] V. S. Borkar and S. K. Mitter, LQG control with communication constraints" in Kailath Festschrift, pp. 365–373, Kluwer Academic Publishers, Boston 1997.

[23] P. Billingsley, Convergence of Probability Measures, New York, NY, John Wiley, 1968

[24] J.H. Braslavsky, R.H. Middleton and J.S. Freudenberg, Feedback stabilization over signal-to-noise ratio constrained channels, IEEE Trans Automatic Control, V. 52, pp. 1391–1403, 2007.

[25] R. Brockett and D. Liberzon, Quantized feedback stabilization of linear systems, IEEE Transactions on Automatic Control, V. 45, pp. 1279-1289, July 2000.

[26] F. Bullo, J. Cortes, and S. Martinez, Distributed Control of Robotic Networks, Applied Mathematics Series, Princeton University Press, 2009.

[27] G. Casalino, F. Davoli, R. Minciardi, P.P. Puliafito and R. Zoppoli, Partially nested information structures with a common past, IEEE Trans. Automatic Control, V. 29, pp. 846-850, Sept. 1984.

[28] C.D. Charalambous, A. Farhadi, D. Denic and F. Rezaei Robust control over uncertain communication channels, in Proc. IEEE Med. Conf. Control and Automation, pp. 737–742, Cyprus, June 2005.

[29] C.D. Charalambous and F. Rezaei, Stochastic uncertain systems subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games, IEEE Transactions on Automatic Control, V. 52, no. 4, pp 647–663, May 2007.

[30] D. Chatterjee, S. Amin, P. Hokayem, J. Lygeros and S. Sastry, Mean-square boundedness of stochastic networked control systems with bounded control inputs, IEEE Conference on Decision and Control, pp. 4759-4764, 2010.

[31] I. Csiszar and J. Korner, Information Theory: Coding Theorems for Discrete Memoryless Channels, Budapest: Akademiai Kiado, 1981.

[32] G. Como, F. Fagnani, and S. Zampieri, Anytime reliable transmission of real-valued information through digital noisy channels, SIAM Journal on Control and Optimization, V. 48, pp. 3903-3924, April 2010.

[33] J. P. Corfmat and A. S. Morse, Decentralized control of linear multivariable systems, Automatica, V. 11, pp. 479-497, Sep. 1976.

[34] T. M. Cover and J. A. Thomas, Elements of Information Theory, Wiley, NY 1991.

[35] I. Csiszár, A class of measures of informativity of observation channels, Period. Math. Hungar., V. 2, pp. 191-213, 1972.

[36] S. Wang and E.J. Davison, On the stabilization of decentralized control. systems, IEEE Trans. Automatic Control, V. 18, pp. 473-478, Oct. 1973.

[37] D. F. Delchamps, Stabilizing a linear system with quantized state feedback, IEEE Trans. Aut. Control, 35(8):916–924, 1990.

[38] R. L. Dobrushin and B. S. Tsybakov, Information transmission with additional noise, IRE Trans. Inform. Theory, V. 18, pp. 293304, 1962.

[39] T. E. Duncan, , On the calculation of mutual information, SIAM Journal of Applied Mathematics, V. 19, pp. 215220, July 1970.

[40] H. Ebeid and T. P. Coleman, Source coding with feedforward using the posterior matching scheme, in Proc. IEEE Int. Symp. on Information Theory (ISIT), Austin, Texas, U.S.A., June 2010.

[41] N. Elia, When Bode meets Shannon: control-oriented feedback communication schemes, IEEE Trans. Aut. Control, 49(9):1477-1488, 2004.

[42] N. Elia and S. K. Mitter, Stabilization of linear systems with limited information, IEEE Trans. Aut. Control, V. 46, N. 9. pp. 1384–1400, 2001.

[43] K. Erciyes, O. Dagdeviren, D. Cokuslu and D. Ozsoyeller, Graphy theoretic clustering algorithms in mobile ad hoc and wireless sensor networks, survey, Appl. Comput. Math. 6, no.2, pp.162-180, 2007.

[44] F. Fagnani and S. Zampieri, Stability analysis and synthesis for scalar linear systems with a quantized feedback, IEEE Trans. Aut. Control, 48(9):1569–1584, 2003.

[45] J. A. Fax and R. M. Murray, Information flow and cooperative control of vehicle formations, IEEE Trans. Automatic Control, V. 49, pp. 1465–1476, Sept. 2004.

[46] T. Fine, Optimum mean-square quantization of a noisy input, IEEE Trans. on Inform. Theory, V. 11, pp. 293294, Apr. 1965.

[47] B. J. Frey, Graphical Models for Machine Learning and Digital Communication, Cambridge, MA: MIT Press, 1998.

[48] R. G. Gallager, Information Theory and Reliable Communication, New York: John Wiley and Sons, 1968.

[49] M. Gastpar, B. Rimoldi, and M. Vetterli, To code, or not to code: Lossy source-channel communication revisited, IEEE Trans. Inform. Theory, V. 49, pp. 11471158, May 2003.

[50] Z.Gong and M. Aldeen, Stabilization of decentralized control systems, Journal of Mathematical Systems, Estimation and Control, V. 7, pp.1-16, 1997.

[51] D. J. Goodman and A. Gersho, Theory of an adaptive quantizer, IEEE Transactions on Communications, 22: 1037- 1045, Aug. 1974.

[52] G. Goodwin and D. Quevado, Finite alphabet control and estimation, International Journal of Control, Automation, and Systems, V. 1, pp. 412-430, 2003.

[53] S. Graf and H. Luschgy, Quantization for probability measures in the Prohorov metric, Theory Probab. Appl., V. 53, pp. 216-241, 2009.

[54] R. M. Gray and T. Hashimoto, A Note on Rate-Distortion Functions for Nonstationary Gaussian Autoregressive Processes, IEEE Trans. Information Theory, V. 54, pp. 1319-1322, March 2008.

[55] R. M. Gray and F. Saadat, Block source coding theory for asymptotically mean stationary sources, IEEE Transactions on Information Theory, pp. 64-67, January 1984.

[56] R. M. Gray, Information rates of autoregressive processes, IEEE Trans. Information Theory, V. 16, pp. 412-421, July 1970.

[57] R. M. Gray and D. L. Neuhoff, Quantization, IEEE Trans. Information Theory, V. 44, pp. 23252383, Oct. 1998.

[58] P. Grover and A. Sahai, , Implicit and explicit communication in decentralized control, Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, September 2010.

[59] A. Gurt and G. Nair Internal stability of dynamic quantised control for stochastic linear plants, Automatica, volume 45, pp. 1387-1396, June 2009.

[60] V. Gupta, N. C. Martins and J. S. Baras, Optimal Output Feedback Control Using Two Remote Sensors Over Erasure Channels, IEEE Transactions on Automatic Control, V. 54, pp. 1463 - 1476, July 2009.

[61] A. György and T. Linder, Optimal entropy-constrained scalar quantization of a uniform source, IEEE Trans. on Inform. Theory, V. 46, pp. 2704 2711, Nov. 2000.

[62] A. György, T. Linder and G. Lugosi, Tracking the best quantizer, IEEE Trans. Inform. Theory, V. 54, pp. 1604,1625, April. 2008.

[63] R. van Handel, The stability of conditional Markov processes and Markov chains in random environments, Ann. Probab., V. 37, pp. 18761925, 2009.

[64] T. Hashimoto and S. Arimoto, On the rate-distortion function for the nonstationary Gaussian autoregressive process, IEEE Trans. Inform. Theory, V. 26, pp. 478-480, 1980.

[65] O. Hernandez-Lerma, J. Gonzales-Hernandez, and R. R. Lopez-Martinez, Constrained average cost Markov control processes in Borel spaces, SIAM J. Control and Optimization, V. 42, pp. 442-468, 2003.

[66] O. Hernandez-Lerma, J. B. Lasserre, Discrete-Time Markov Control Processes: Basic Optimality Criteria, Springer, New York, 1996.

[67] Y. C. Ho, Team decision theory and information structures, Proc. IEEE, V. 68, pp. 644654, 1980.

[68] Y. C. Ho and K. C. Chu Team decision theory and information structures in optimal control problems - Part I, IEEE Trans. Automatic Control, 17, pp. 15 - 22, Feb. 1972.

[69] M. Huang and S. Dey, Dynamic quantizer design for hidden Markov state estimation via multiple sensors with fusion center feedback, IEEE Tran. Signal Processing 54: 2887-2896, 2006.

[70] O.C. Imer and T. Başar, Optimal estimation with scheduled measurements, Applied and Computational Math., 4(2):92-101, 2005.

[71] O. C. Imer, S. Yüksel, and T. Başar, Optimal control of LTI systems over communication networks, Automatica, 42(9):1429–1440, 2006.

[72] H. Ishii and T. Başar, Remote control of LTI systems over networks with state quantization, Systems & Control Letters, 54(1):15–32, 2005.

[73] H. Ishii and B. A. Francis, Limited Data Rate in Control Systems with Networks, Lecture Notes in Control and Information Sciences, Vol. 275, Springer, Berlin, 2002.

[74] A. P. Johnston and S. Yüksel, Stochastic Stabilization of Partially Observed and Multi-Sensor Systems Driven by Gaussian Noise under Fixed-Rate Information Constraints, Queen's University Internal Report, 2011.

[75] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Transaction of the ASME Journal of Basic Engineering, pp. 35-45, March 1960.

[76] J. C. Kieffer and J. G. Dunham, On a type of stochastic stability for a class of encoding schemes, IEEE Transactions on Information Theory, 29: 793-797, November 1983.

[77] P. P. Khargonekar and A. B. Özgüler, Decentralized control and periodic feedback, IEEE Transactions on Automatic Control, V. 39, pp. 877-882, April 1994.

[78] T. Keviczky, F. Borrelli and G. J. Balas, Decentralized receding horizon control for large scale dynamically decoupled systems, Automatica, December 2006, Vol. 42, No. 12, pp. 2105-2115.

[79] H. Kobayashi, H. Hanafusa and T. Yoshikawa, Controllability under decentralized information structure, IEEE Trans. Automatic Control V. 23, pp. 182-188, Apr. 1978.

[80] B. Kurtaran, Corrections and extensions to Decentralized control with delayed sharing information pattern, IEEE Trans. Automatic Control, vol 24, pp. 656-657, Aug. 1979.

[81] D. Liberzon, On stabilization of linear systems with limited information, IEEE Trans. Aut. Control, 48(2):304–307, 2003.

[82] T. Linder and G. Lugosi, A zero-delay sequential scheme for losssy coding of individual sequences, IEEE Trans. on Information Theory, V. 47, pp. 2533-2538, Sep. 2001.

[83] T. Linder and R. Zamir, Causal coding of stationary sources and individual sequences with high resolution, IEEE Trans. Inform. Theory, V. 52, pp. 662-680, Feb. 2006.

[84] G. Lipsa and N. C. Martins, Certifying the optimality of a distributed state estimation system via majorization theory, ISR Technical Report TR 2009-19.

[85] A. Mahajan, A. Nayyar and D. Teneketzis, Identifying tractable decentralized control problems on the basis of information structure, in Proc. Annual Allerton Conference on Communications, Control and Computing, Allerton, IL, September 2008.

[86] A. Mahajan, Sequential decomposition of sequential teams: applications to real-time communication and networked control systems, Ph.D. Dissertation, University of Michigan, Ann Arbor, September 2008.

[87] A. Mahajan and D. Teneketzis, On the design of globally optimal communication strategies for real-time noisy communication with noisy feedback, IEEE Journal on Special Areas in Communications, V. 28, pp. 580-595, May 2008.

[88] A. Mahajan and S. Tatkinda, An axiomatic approach for simplification of sequential teams, IEEE Trans. Automatic Control, submitted.

[89] A. Mahajan and D. Teneketzis, Optimal design of sequential real-time communication systems, IEEE Transactions on Inform. Theory, V. 55, pp. 5317-5338, November 2009.

[90] A. Mahajan and D. Teneketzis, On the design of globally optimal communication strategies for real-time noisy communication with noisy feedback, IEEE Journal on Selected Areas in Comm., 26:580–595, May, 2008.

[91] A. Mahajan and S. Yüksel, Measure and cost dependent properties of information structures, Proc. IEEE American Control Conference, July 2010, Baltimore, MD, USA.

[92] N. C. Martins, M. A. Dahleh and N. Elia, Feedback stabilization of uncertain systems in the presence of a direct link, IEEE Trans. Aut. Control, V. 51, N. 3, pp. 438–447, 2006.

[93] N. C. Martins and M. A. Dahleh, Feedback Control in the Presence of Noisy Channels: Bode-Like Fundamental Limitations of Performance, IEEE Transactions on Automatic Control, V. 53, pp. 1604 - 1615, Aug. 2008.

[94] N. C. Martins, M. A. Dahleh and John C. Doyle, Fundamental Limitations of Disturbance Attenuation in the Presence of Side Information, IEEE Transactions on Automatic Control, V. 52, pp. 56 - 66, Jan. 2007

[95] N. C. Martins, Finite Gain lp Stability Requires Analog Control, Systems and Control Letters, V. 55, pp. 949-954, November 2006.

[96] A. S. Matveev and A. V. Savkin Stabilization of multi-sensor networked control systems with communication constraints, in Proc. Asian Control Conference, Melbourne, Australia, July 2004.

[97] A. S. Matveev and A. V. Savkin Decentralized stabilization of linear systems via limited capacity communication networks, Proc. of the 44th IEEE CDC and ECC 2005, pp. 11551161, Seville, Spain, Dec. 2005.

[98] A. S. Matveev and A. V. Savkin Decentralized stabilization of networked systems under data-rate constraints" in Proc. International Federation of Automatic Control, Seoul, Korea, July 2008.

[99] A. S. Matveev and A. V. Savkin, Estimation and Control over Communication Networks, Birkhäuser Boston, 2008.

[100] S. P. Meyn and R. Tweedie, Markov Chains and Stochastic Stability, Springer Verlag, London, 1993.

[101] S. P. Meyn and R. Tweedie, Stability of Markovian processes I: Criteria for discrete-time chains, Advances in Applied Probability, 24: 542-574, 1992.

[102] S. P. Meyn and R. Tweedie, State-dependent criteria for convergence of Markov chains, Ann. Appl. Prob., V. 4, pp. 149–168, 1994.

[103] P. Minero, M. Franceschetti, S. Dey and G. Nair, Data rate theorem for stabilization over time-varying feedback channels, IEEE Trans. Aut. Control, V. 54, N. 2, pp. 243-255, 2009.

[104] Y. Kaspi and N. Merhav, Structure theorem for real-time variable-rate lossy source encoders and memory-limited decoders with side information, in Proc. IEEE Int. Symp. on Information Theory (ISIT), Austin, Texas, U.S.A., June 2010.

[105] W. Naiqi, L. Renhou and H. Baosheng Decentralized controllability and observability, and decentralized stabilization, in Proc. IEEE Internation Conference on Systems, Man, and Cybernetics, pp. 598-601, 8-12 Aug 1988.

[106] G. Nair, F. Fagnani, S. Zampieri, J.R. Evans, Feedback control under data constraints: an overview, Proceedings of the IEEE, pp.108-137, 2007.

[107] G. N. Nair, R. J. Evans and P. E. Caines, Stabilising decentralised linear systems under data rate constraints, in Proc. 43rd IEEE CDC, pp. 3992 - 3997, The Bahamas, December 2004.

[108] G. N. Nair and R. J. Evans, Stabilizability of stochastic linear systems with finite feedback data rates, SIAM Journal on Control and Optimization, 43: 413 - 436, July 2004.

[109] G. N. Nair and R. J. Evans, Cooperative networked stabilisability of linear systems with measurement noise, Proc. 15th IEEE Mediterranean Conf. Control and Automation, Athens, Greece, June 2007.

[110] A. Nayyar and D. Teneketzis, On the structure of real-time encoders and decoders in a multi-terminal communication system, submitted to IEEE Transactions on Inform. Theory, 2009 (available on arXiv).

[111] D. L. Neuhoff and R. K. Gilbert, Causal source codes", IEEE Trans. Inform. Theory, 28:701-713, September 1982.

[112] A. Nayyar, A. Mahajan and D. Teneketzis, Optimal control strategies in delayed sharing information structures , IEEE Trans. Automatic Control, to appear.

[113] J. Ooi, S. Verbout, J. Ludwig and G. Wornell, A separation theorem for periodic sharing information patterns in decentralized control, IEEE Trans. Automatic Control, V. 42, pp. 1546 1550, Nov 1997.

[114] A. Orlitsky and J. R. Roche, Coding for computing, IEEE Trans. Information Theory, V. 47, pp. 903 - 917, March 2001.

[115] B. Oksendal, Stochastic Differential Equations, Berlin: Springer, 2003.

[116] J. K. Omura Expurgated bounds, Bhattacharyya distance, and rate distortion function, Inform. and Control, 24:358–383, 1974.

[117] A. Orlitsky and J. R. Roche, Coding for computing, IEEE Trans. Information Theory, V. 47, pp. 903 - 917, March 2001.

[118] U. Özgüner and E. J. Davison, Sampling and decentralized fixed modes, IEEE American Control Conference, pp. 257-262, 1985.

[119] R. Ostrovsky, Y. Rabani, and L. Schulman, Error-correcting codes for automatic control, IEEE Trans. Inform. Theory, V. 55(7):2931-2941, 2009.

[120] S. S. Pradhan and K. Ramchandran, Distributed source coding using syndromes (DISCUS): Design and construction, IEEE Trans. Information Theory, V. 49, pp. 626 - 643, March 2003.

[121] R. Radner, Team decision problems, Annals of Mathematical Statistics, V. 33, pp. 857 - 881, 1962

[122] A. Rantzer, Linear Quadratic Team Theory Revisited, Proc. IEEE American Control Conference, Minneapolis, June 2006.

[123] M. Rotkowitz and S Lall, A characterization of convex problems in decentralized Control, IEEE Trans. Automatic Control, vol 51, pp. 274-286, February 2006.

[124] A. Sahai and S. 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. Inform. Theory, V. 52, N. 8, pp. 3369–3395, 2006.

[125] A. Sahai, Anytime Information Theory, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.

[126] S. V. Sarma, M.A. Dahleh, and S. Salapaka Synthesis of efficient time-varying bit-allocation strategies maintaining input-ouput stability, in Proc. Allerton Conf., October 2004.

[127] A. Sahai and S. 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. Inform. Theory, V. 52, pp. 3369–3395, Aug. 2006.

[128] N. Sandell, P. Varaiya, M. Athans, and M. Safonov, Survey of decentralized control methods for large scale systems, IEEE Trans. Automatic Cotrol, V. 23, pp. 108- 128, April 1978.

[129] N. Sandell and M. Athans, Solution of some nonclassical LQG stochastic decision problems, IEEE Trans. Automatic Control, vol 19, pp. 108-116, 1974.

[130] A. N. Shiryaev, On Markov sufficient statistics in nonadditive Bayes problems of sequential analysis, Theory Probab. Appl., V. 9, pp. 604-618, 1964.

[131] J. D. Slepian and J. K. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Information Theory, 19:471–480, July 1973

[132] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla and S. S. Sastry Foundations of control and estimation over lossy networks, Proc. of the IEEE, V. 95, N.1, pp. 163–187, 2007.

[133] T. Simsek and P. Varaiya, Noisy data-rate limited estimation: Renewal codes, in Proc. IEEE Conf. Decision and Control, pp. 3149–3154, Dec. 2003.

[134] A. J. Viterbi and J. K. Omura, Principles of Digital Communication and Coding, McGraw-Hill, Inc. New York, 1979.

[135] V. Poor, An introduction to signal detection and estimation, Springer, 1994.

[136] C. Striebel, Sufficient statistics in the optimum control of stochastic systems, J. Math. Anal. Appl., V. 12, pp. 576-592, 1965.

[137] N.R. Sandell, Jr., P. Variaya, M. Athans and M.G. Safonov, Survey of decentralized control methods for large scale systems, IEEE Trans. Automatic Control, V. 23, pp. 108-128 April 1978.

[138] S. Tatikonda, Some scaling properties of large distributed control systems, Proc. 42nd IEEE CDC, Maui, Hawaii, pp. 3142 - 3147, December 2003.

[139] S. Tatikonda, Cooperative control under communication constraints, Proc. IEEE Information Theory Workshop (ITW), May 2008.

[140] S. Tatikonda, Control Under Communications Constraints, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[141] S. Tatikonda and S. Mitter, Control under communication constraints, IEEE Trans. Aut. Control, 49(7):1056-1068, 2004.

[142] S. Tatikonda, A. Sahai, and S. Mitter, Stochastic linear control over a communication channels, IEEE Trans. Aut. Control, V. 49, pp. 1549-1561, Sept. 2004.

[143] S. Tatikonda and M. I. Jordan, Loopy belief propagation and Gibbs measures, in Proc. Uncertainty in Artificial Intelligence, V. 18, pp. 493-500, Aug. 2002.

[144] D. Teneketzis, On information structures and nonsequential stochastic control, CWI Quarterly, V. 9, pp. 241-260, 1996.

[145] D. Teneketzis, On the structure of optimal real-time encoders and decoders in noisy communication, IEEE Transactions on Inform. Theory, V. 52, pp. 4017-4035, September 2006.

[146] J. N. Tsitsiklis and M. Athans, Convergence and asymptotic agreement in distributed decision problems, IEEE Trans Automatic Control, Vol. 29, pp. 42-50, Jan. 1984.

[147] P. G. Voulgaris, A convex characterization of classes of problems in control with specific interaction and communication structures, Proc. IEEE American Control Conference, pp. 3128-3133, Arlington, VA, June 2001.

[148] R. Venkataramanan and S. S. Pradhan, Source coding with feedforward: Rate-distortion theorems and error exponents for a general source, IEEE Transactions on Information Theory, V. 53, pp. 2154-2179, June 2007.

[149] J. C. Walrand and P. Varaiya, Optimal causal coding-decoding problems, IEEE Trans. Inform. Theory, V. 19, pp. 814-820, November 1983.

[150] S. H. Wang, Stabilization of decentralized control systems via time-varying controllers, IEEE Trans. Automatic Control, V. 27, pp. 741 - 744, June 1982.

[151] T. Weissman and N. Merhav, On causal source codes with side information, IEEE Transactions on Information Theory, V. 51, pp. 40034013, November 2005.

[152] J. L. Willems, Time-varying feedback for the stabilization of fixed modes in decentralized control systems, Automatica, V. 25, pp. 127-131, 1989,

[153] H. S. Witsenhausen, The intrinsic model for discrete stochastic control: Some open problems," in Control Theory, Numerical Methods and Computer System Modelling, Lecture Notes in Economics and Mathematical Systems, A. Bensoussan and J. L. Lions, Eds. Springer Verlag, 1975, V. 107, pp. 322–335.

[154] H. S. Witsenhausen, Indirect rate-distortion problems, IEEE Trans. Inform. Theory, V. 26, pp. 518521, Sept. 1980.

[155] H. S. Witsenhausen, On the structure of real-time source coders, Bell Syst. Tech. J., 58:1437-1451, July/August 1979.

[156] H. S. Witsenhausen, A counterexample in stochastic optimum control, SIAM Journal on Control and Optimization, V. 6, pp. 131-147, 1968.

[157] H. S. Witsenhausen, Equivalent stochastic control problems, Mathematics of Control, Signals, and Systems, V. 1, pp. 3-11, Springer-Verlag, 1988.

[158] H. S. Witsenhausen, On information structures, feedback and causality, SIAM J. Control, V. 9, pp. 149160, May 1971.

[159] W. S. Wong, Control communication complexity of distributed control systems, SIAM Journal on Control and Optimization, V. 48, pp. 1722-1742, May 2009.

[160] P. W. Wong and R. M. Gray, Sigma-delta modulation with i.i.d. Gaussian inputs, IEEE Transactions on Information Theory, Vol. IT-36, pp. 784-778, July 1990.

[161] W. S. Wong and R. W. Brockett, Systems with finite communication bandwidth constraints - part II: Stabilization with limited information feedback, IEEE Trans. Automatic Control, V. 42, pp. 1294 - 1299, September 1997.

[162] A.C.-C. Yao, Some complexity questions related to distributive computing, Proc. of the 11th Annual ACM Symposium on Theory of Computing, 1979.

[163] T. Yoshikawa, Decomposition of dynamic team decision problems, IEEE Trans. Automatic Control, V. 23, pp. 627-632, August 1978.

[164] K. You and L. Xie, Minimum data rate for mean square stabilization of discrete LTI systems over lossy channels IEEE Trans. Aut. Control, to appear.

[165] S. Yüksel, Optimal LQG coding and control over communication channels: An existence and an infeasibility result, in Proc. IEEE American Control Conf., Seattle, Washington, June 2008.

[166] S. Yüksel, Stochastic stability and ergodicity properties of linear systems controlled over noisy channels, *Proc. IEEE Decision and Control Conference*, December 2011.

[167] S. Yüksel, A random time stochastic drift result and application to stochastic stabilization over noisy channels, in Proc. Annual Allerton Conference, IL, October 2009.

[168] S. Yüksel, Stochastic stability of adaptive quantizers for Markov sources, Proc. IEEE ISIT, Seoul, S. Korea, June 2009.

[169] S. Yüksel, On optimal causal coding of partially observed Markov Sources under classical and non-classical information structures, Proc. IEEE ISIT, Austin, TX, June 2010.

[170] S. Yüksel, On optimal causal coding of partially observed Markov Sources in single and multi-terminal settings, available on arXiv.

[171] S. Yüksel, Stochastic stabilization of noisy linear systems with fixed-rate limited feedback, IEEE Trans. Automatic Control, vol. 55, pp. 2847-2853, December 2010.

[172] S. Yüksel, Stochastic nestedness and the belief sharing information pattern", IEEE Trans. Automatic Control, V. 55, pp. 2773-2786, Dec. 2009.

[173] S. Yüksel and T. Başar, Information theoretic study of the signaling problem in decentralized stabilization of noisy linear systems, Proc. IEEE CDC, New Orleans, Dec. 2007.

[174] S. Yüksel and T. Başar, On the absence of rate loss in decentralized sensor and controller structure for asymptotic stability," Proc. IEEE American Control Conf., Minneapolis, MN, June 2006.

[175] S. Yüksel and T. Başar, Communication constraints for decentralized stabilizability with time-invariant policies, IEEE Trans. Automatic Control, V. 52, pp. 1060-1066, June 2007.

[176] S. Yüksel and T. Başar, Optimal signaling policies for decentralized multi-controller stabilizability over communication channels, IEEE Trans. Automatic Control, V. 52, pp. 1969-1974, October 2007.

[177] S. Yüksel and T. Başar, Minimum rate coding for LTI systems over noiseless channels, IEEE Trans. Aut. Control, V. 51, N. 12, pp. 1878–1887, 2006.

[178] S. Yüksel, T. Başar and S. P. Meyn Optimal causal quantization of Markov Sources with distortion constraints, Proc. Information Theory and Applications Workshop (ITA), San Diego, CA, U.S.A., Jan. 2008.

[179] S. Yüksel and T. Linder, Optimization and convergence of observation channels in stochastic control, submitted (available on arXiv).

[180] S. Yüksel and S. P. Meyn, Random-time, state-dependent stochastic drift for Markov Chains and application to stochastic stabilization over erasure channels, IEEE Trans. Automatic Control, under review.

[181] S. Yüksel and T. Başar, Control over noisy forward and reverse channels, IEEE Transactions on Automatic Control, V. 56, pp. 1014-1029, May 2011.