Article

Tracking an Auto-Regressive Process with Limited Communication per Unit Time †

Rooji Jinan [1,*]📵, **Parimal Parag** [2]📵 and **Himanshu Tyagi** [2]

[1] Robert Bosch Centre for Cyber-Physical Systems, Indian Institute of Science, Bangalore 560012, India
[2] Department of ECE, Indian Institute of Science, Bangalore 560012, India; parimal@iisc.ac.in (P.P.); htyagi@iisc.ac.in (H.T.)
[*] Correspondence: roojijinan@iisc.ac.in
[†] This paper is an extended version of our paper published in the Proceeding of the 2020 IEEE International Symposium on Information Theory (ISIT), June 2020 titled "Tracking an Auto-Regressive Process with Limited Communication".

Abstract: Samples from a high-dimensional first-order auto-regressive process generated by an independently and identically distributed random innovation sequence are observed by a sender which can communicate only finitely many bits per unit time to a receiver. The receiver seeks to form an estimate of the process value at every time instant in real-time. We consider a time-slotted communication model in a slow-sampling regime where multiple communication slots occur between two sampling instants. We propose a *successive update* scheme which uses communication between sampling instants to refine estimates of the latest sample and study the following question: Is it better to collect communication of multiple slots to send better refined estimates, making the receiver wait more for every refinement, or to be *fast but loose* and send new information in every communication opportunity? We show that the fast but loose successive update scheme with ideal spherical codes is universally optimal asymptotically for a large dimension. However, most practical quantization codes for fixed dimensions do not meet the ideal performance required for this optimality, and they typically will have a bias in the form of a fixed additive error. Interestingly, our analysis shows that the fast but loose scheme is not an optimal choice in the presence of such errors, and a judiciously chosen frequency of updates outperforms it.

Keywords: auto-regressive process; quantization; successive update scheme

1. Introduction

We consider the setting of real-time decision systems based on remotely sensed observations. In this setting, the decision maker needs to track the remote observations with high precision and in a timely manner. These are competing requirements, since high precision tracking will require larger number of bits to be communicated, resulting in larger transmission delay and increased staleness of information. Towards this larger goal, we study the following problem.

Consider a discrete time first-order auto-regressive (AR[1]) process $X_t \in \mathbb{R}^n$, $t \geq 0$. A sensor draws a sample from this process, periodically once every s time-slots. In each of these time-slots, the sensor can send nR bits to a center. The center seeks to form an estimate $\hat{X}_t$ of X_t at time t, with small mean square error (MSE). Specifically, we are interested in minimizing the time-averaged error $\sum_{t=1}^{T} \mathbb{E}\|X_t - \hat{X}_t\|_2^2 / T$ to enable timely and accurate tracking of X_t.

We propose and study a *successive update* scheme where the encoder computes the error in the estimate of the latest sample at the decoder and sends its quantized value to the decoder. The decoder adds this value to its previous estimate to update the estimate of the latest sample, and uses it to estimate the current value using a linear predictor. We instantiate this scheme with a general gain-shape quantizer for error-quantization.

Note that we can send this update several times between two sampling instances. In particular, our interest is in comparing a *fast but loose* scheme where an update is sent every slot against a scheme with slower update in every p communication slots. The latter allows the encoder to use more bits for the update, but the decoder will need to wait longer. We consider a class $\mathbb{X}_n$ of discrete time AR[1] processes generated by an independently and identically distributed (*i.i.d*) random innovation sequence such that the fourth moment of the process is bounded. Within this class, we show that the fast but loose successive update scheme, used with an appropriately selected quantizer, is universally optimal for all possible distributions of the innovation sequence when the number of dimensions grows asymptotically large.

To show this optimality, we use a random construction for the quantizer, based on the spherical code given in [1,2]. Roughly speaking, this ideal quantizer Q yields

$$\mathbb{E}\|y - Q(y)\|_2^2 \leqslant \|y\|_2^2 2^{-2R},$$

for every n dimensional vector y uniformly bounded. However, in practice, at finite n, such quantizers need not exist. Most practical vector quantizers have an extra additive error, *i.e.*, the error bound takes the form

$$\mathbb{E}\|y - Q(y)\|_2^2 \leqslant \|y\|_2^2 \theta + n\varepsilon^2$$

where θ and ϵ are quantizer parameters that varies with the choice of quantizer. We present our analysis for such general quantizers. Interestingly, for such a quantizer (which is all we have at a finite n), the optimal choice of p can differ from 1. Our analysis provides a theoretically sound guideline for choosing the frequency of updates $1/p$ for practical quantizers.

Our work relates to a large body of literature ranging from real-time compression to control and estimation over networks. The structure of real-time encoders for source coding has been studied in [3–9]. The general structure of real-time encoders for Markov sources is studied for communication over error-free channels in [3] and over noisy channels in [4,5]. The authors in [3] consider the communication delays similar to our setting. However, the delayed distortion criterion considered in [3] is different from the instantaneous distortion in our work. From these works, we see that optimal encoder output of a kth order Markov source at any instant would depend only on the k latest symbols and the present state of decoder memory. A similar structural result for the optimal encoders and decoders which are restricted to be causal is given in [6]. Furthermore, structural results in the context of optimal zero-delay coding of correlated sources are available in [7–9]. Some of these results can be extended to the case of finite delay decoding. However, as we need to track the process in real time, we have to produce an estimate for the latest sample at every instant even though the current information available at the decoder is stale and corresponds to past samples. This is quite different from the case of delayed decoding. Hence, the setup in all these works is different from the problem we consider and the results do not extend to our problem. Another related area which has been studied in literature is that of remote source coding of noisy sources [10–12] with an optimal encoder-decoder structure discussed in [11]. Further, [13] examines the optimal sampling strategy for remote reconstruction of a bandlimited signal from noisy versions of the source. Studies about remote reconstruction of a stationary process from its noisy/noiseless samples can be found in [14–17] as well. However, in our setting, the sampling is fixed and moreover each transmission in our system incurs a delay that depends on the encoding rate. Hence, our problem does not directly fit into any of these frameworks.

The problems of remote estimation under communication constraints of various kinds have been studied in [18–23]. This line of work proposes several Kalman-like recursive estimation algorithms and evaluates their performances. In a related thread, [24–26] study remote estimation under communication constraints and other related constraints using tools from dynamic programming and stochastic control. However, in all these works the

role of channel delay is slightly different from that in our setting. Furthermore, the specific problem of choice of quantizer we consider has not been looked at. More recently, [27] studied remote estimation of Wiener process for channels with random delays and proves that the optimal sampling policy is a threshold based policy. This work, like some of the works cited above, assumes real-valued transmissions and does not take into account any quantization effects.

In more information theoretic settings, sequential coding for individual sequences under delay constraints was studied in [28–31]. Closer to our work, the causal (nonanticipatory) rate-distortion function for a stochastic process goes back to early works [32,33]. Recent works [34–36] consider the specific case of auto-regressive and Gauss–Markov processes and use the general formula in these early works to establish asymptotic optimality of simpler information structure for the encoders (optimal decoder structure is straightforward). Further, the system model in these works slightly differs from ours as the information transmission in our setting suffers a delay due to the channel rate constraint. We note that related formulations have been studied for simple settings of two or three iterations in [37,38] where interesting encoder structures for using previous communication for next sample as well emerge. Although some of these works propose specific optimal schemes as well, the key results in this line of research provide an expression for the rate-distortion function as an optimization problem, solving which will provide guidelines for a concrete scheme. In contrast, motivated by problems of estimation over an erasure channel, ref. [39] provides an asymptotically optimal scheme that roughly uses Gaussian codebooks to quantize the innovation errors between the encoder's observation and the decoder's estimation. Our work is closest to [39] and our encoding scheme shares similarities with the predictive Differential Pulse Code Modulation (DPCM) based scheme employed in [39]. However, our work differs in an important aspect from all the works mentioned in this research thread: we take transmission delays into account. In particular, in our formulation the estimation at the decoder happens in real time in spite of the communication delays. Note that the rate constraint in the channel causes a delay in the reception of information at the decoder. Nevertheless, the decoder must provide an estimate of the current state of the process at every time instant, and a longer codeword will result in a longer delay for the decoder to get complete information.

Nonetheless, our converse bound is derived using similar methods as [39]. Even the achievability part of our proof draws from [39], but there is a technical caveat. Note that after the first round of quantization, the error vector need not be Gaussian, and the analysis in [39] can only be applied after showing a closeness of the error vector distribution to Gaussian in the Wasserstein distance of order 2. While the original proof [39] overlooks this technical point, this gap can be filled using a recent result from [40] if spherical codes are used. However, we follow an alternative approach and show a direct analysis using vector quantizers.

In addition, there is a large body of work on control problems over rate-limited communication channels (cf. [41–50]). This line of work implicitly requires handling of communication delays in construction of estimators. However, the simple formulation we have seems to be missing and the results in this long line of work do not resolve the questions we raise.

Our main contributions in this paper are as follows: we present an encoder structure which we show to be optimal asymptotically in the dimension of the observation. Specifically, we propose to send successive updates that refines the estimates of the latest sample at the decoder. It is important to note that we quantize the estimation error at the decoder, instead of quantizing the innovation sequence formed at the encoder. Although the optimal MMSE decoder involves taking conditional expectation, we use a simple decoder which uses a simple linear structure. Yet, we show that this decoder is asymptotically optimal. Then, we instantiate this general scheme with spherical codes for quantizers to obtain a universal scheme. In particular, we consider general gain-shape quantizers and develop a framework to analyze their performance. One interesting result we present shows that the

tradeoff between the accuracy and frequency of the updates must be carefully balanced based on the "bias" (additive error) in the quantizer used.

We present our problem formulation in the next section. Section 3 presents a discussion on the our achievability scheme followed by the main results in the subsequent section. Section 5 provides a detailed analysis of our scheme, which we further build-on in Section 6 to get our asymptotic achievability results. We prove our converse bound in Section 7 and conclude with a discussion on extensions of our result in the final section.

2. Problem Formulation

We begin by providing a formal description of our problem. Figure 1 illustrates the communication model adopted in this work. Different components of the model are presented in separate sections. Throughout the remainder of this paper, the set of real numbers is denoted by $\mathbb{R}$, the set of positive reals is denoted by $\mathbb{R}_+$, the n-dimensional Euclidean space is denoted by $\mathbb{R}^n$ and the associated Euclidean norm by $\|\cdot\|_2$, the set of positive integers is denoted by $\mathbb{N}$, the set of non-negative integers is denoted by $\mathbb{Z}_+$, the set of continuous positive integers until m is denoted by $[m] \triangleq \{1, \ldots, m\}$, and an identity matrix of size $n \times n$ is denoted by I_n.
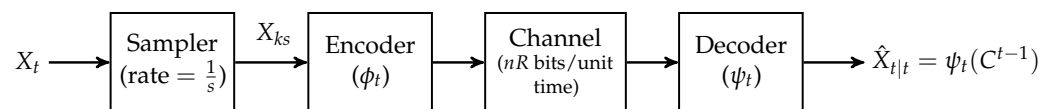

Figure 1. Communication model.

2.1. Observed Random Process and Its Sampling

For $\alpha \in (0, 1)$, we consider a discrete time auto-regressive process of order 1 (AR[1] process) in $\mathbb{R}^n$,

$$X_t = \alpha X_{t-1} + \xi_t, \quad t \geqslant 0, \tag{1}$$

where $(\xi_t \in \mathbb{R}^n, t \geqslant 1)$ is an *i.i.d.* random sequence with zero mean and covariance matrix $\sigma^2(1 - \alpha^2)I_n$. For simplicity, we assume that $X_0 \in \mathbb{R}^n$ is a zero mean random variable with covariance matrix $\sigma^2 I_n$. This choice of covariance matrices for ξ_t and X_0 ensures that the variance of $X_t \in \mathbb{R}^n$ is $\sigma^2 I_n$ for all $t \geqslant 0$. Although this assumption makes the analysis more tractable, the dependence of the covariance matrix of ξ_t on the process parameter α is not crucial for the analysis.

In addition, we assume that $\|X_t\|_2$ has a bounded fourth moment at all times $t \geqslant 0$. Specifically, let $\kappa > 0$ satisfy

$$\sup_{k \in \mathbb{Z}_+} \frac{1}{n} \sqrt{\mathbb{E} \|X_k\|_2^4} \leqslant \kappa.$$

It is clear that $X = (X_t \in \mathbb{R}^n, t \geqslant 0)$ is a Markov process. We denote the set of processes X satisfying the assumptions above by $\mathbb{X}_n$ and the class of all such processes for different choices of dimension n as $\mathbb{X}$.

This discrete time process is subsampled periodically at *sampling frequency* $1/s$, for some $s \in \mathbb{N}$, to obtain samples $(X_{ks} \in \mathbb{R}^n, k \geqslant 0)$.

2.2. Encoder Description

The sampled process $(X_{ks}, k \geqslant 0)$ is passed to an *encoder* which converts it to a bit stream. The encoder operates in *real-time* and sends nRs bits between any two sampling instants. Specifically, the encoder is given by a sequence of mappings $(\phi_t)_{t \geqslant 0}$, where the mapping at any discrete time $t = ks$ is denoted by

$$\phi_t : \mathbb{R}^{n(k+1)} \to \{0, 1\}^{nRs}.$$

The encoder output at time $t = ks$ is denoted by the codeword $C_{ks} \triangleq \phi_k(X_0, X_s, \ldots, X_{ks})$. We denote this codeword by an s-length sequence of binary strings $C_{ks} = (C_{ks,0}, \ldots, C_{ks,s-1})$,

where each term $C_{ks,i}$ takes values in $\{0,1\}^{nR}$. For time ks and $0 \leqslant i \leqslant s - 1$, we can view the binary string $C_{ks,i}$ as the communication sent at time $ks + i$. We elaborate on the communication channel next.

2.3. Communication Channel

The output bit-stream of the encoder is sent to the receiver via an error-free communication channel. Specifically, we assume slotted transmission with synchronization where in each slot the transmitter sends nR bits of communication error-free. That is, we are allowed to send R bits per dimension, per time slot. Note that there is a delay of 1 time-unit (corresponding to one slot) in transmission of each nR bits. Therefore, the vector $C_{ks,i}$ of nR bits transmitted at time $ks + i$ is received at time instant $ks + i + 1$ for $0 \leqslant i \leqslant s - 1$. Throughout we use the notation $I_k \triangleq \{ks, \ldots, (k+1)s - 1\}$ and $\tilde{I}_k = I_k + 1 = \{ks + 1, \ldots, (k+1)s\}$, respectively, for the set of transmit and receive times for the strings $C_{ks,i}$, $0 \leqslant i \leqslant s - 1$.

2.4. Decoder Description

We describe the operation of the receiver at time $t \in I_k$, for some $k \in \mathbb{N}$, such that $i = t - ks \in \{0, \ldots, s - 1\}$. Upon receiving the codewords $C_s, C_{2s}, \ldots, C_{(k-1)s}$ and the partial codeword $(C_{ks,0}, \ldots, C_{ks,i-1})$ at time $t = ks + i$, the *decoder* estimates the current-state X_t of the process using the estimator mapping

$$\psi_t : \{0,1\}^{nRt} \to \mathbb{R}^n.$$

We denote the overall communication received by the decoder until time instant t by C^{t-1}. The time index $t - 1$ in C^{t-1} corresponds to the transmission time of the codewords, whereby the communication received till time t is denoted by C^{t-1}. Further, we denote by $\hat{X}_{t|t}$ the real-time causal estimate $\psi_t(C^{t-1})$ of X_t formed at the decoder at time t. Thus, the overall real-time causal estimation scheme is described by the mappings $(\phi_t, \psi_t)_{t \geqslant 0}$. It is important to note that the communication available to the decoder at time $t \in I_k$ can only depend on samples X_ℓ up to time $\ell \leqslant ks$. As a convention, we assume that $\hat{X}_{0|0} = 0$.

2.5. Performance Metrics

We call the encoder-decoder mapping sequence $(\phi, \psi) = (\phi_t, \psi_t)_{t \geqslant 0}$ a tracking code of rate R and sampling period s. The tracking error of our tracking code at time t for process X is measured by the mean squared error (MSE) per dimension given by

$$D_t(\phi, \psi, X) \triangleq \frac{1}{n} \mathbb{E} \| X_t - \hat{X}_{t|t} \|_2^2.$$

Our goal is to design (ϕ, ψ) with low average tracking error $\overline{D}_T(\phi, \psi, X)$ given by

$$\overline{D}_T(\phi, \psi, X) \triangleq \frac{1}{T} \sum_{t=0}^{T-1} D_t(\phi, \psi, X).$$

For technical reasons, we restrict to a finite time horizon setting. For the most part, the time horizon T will remain fixed and will be omitted from the notation. Instead of working with the mean-square error, a more convenient but equivalent parameterization for us will be that of accuracy, given by

$$\delta^T(\phi, \psi, X) = 1 - \frac{\overline{D}_T(\phi, \psi, X)}{\sigma^2}.$$

Definition 1 (Maxmin tracking accuracy). *The worst-case tracking accuracy for $\mathbb{X}_n$ attained by a tracking code (ϕ, ψ) is given by*

$$\delta^T(\phi, \psi, \mathbb{X}_n) = \inf_{X \in \mathbb{X}_n} \delta^T(\phi, \psi, X).$$

The maxmin tracking accuracy for $\mathbb{X}_n$ at rate R and sampling period s is given by

$$\delta_n^T(R, s, , \mathbb{X}_n) = \sup_{(\phi, \psi)} \delta^T(\phi, \psi, \mathbb{X}_n),$$

where the supremum is over all tracking codes (ϕ, ψ).

The maxmin tracking accuracy $\delta_n^T(R, s, \mathbb{X}_n)$ is the fundamental quantity of interest for us. Note that it is possible to provide a similar characterization of the performance in terms of MSE instead of accuracy and obtain the exact same result. Recall that n denotes the dimension of the observations in X_t for $X \in \mathbb{X}_n$ and T the time horizon. However, we will only characterize $\delta_n^T(R, s, \mathbb{X}_n)$ asymptotically in n and T. Specifically, we define the asymptotic maxmin tracking accuracy as

$$\delta^*(R, s, \mathbb{X}) = \limsup_{T \to \infty} \limsup_{n \to \infty} \delta_n^T(R, s, \mathbb{X}_n).$$

We will provide a characterization of $\delta^*(R, s, \mathbb{X})$ and present a sequence of tracking codes that attains it. In fact, the tracking code we use is an instantiation of our successive update scheme, which we describe in the next section. It is important to note that our results may not hold if we switch the order of limits above: we need very large codeword lengths depending on a fixed finite time horizon T.

3. The Successive Update Scheme

In this section, we present our main contribution in this paper, namely the *Successive Update* tracking code. Before we describe the scheme completely, we present its different components. In every communication slot, the transmitter gets an opportunity to send nR bits. The transmitter may use it to send any information about a previously seen sample. There are various options for the encoder. For instance, it may use the current slot to send some information about a sample it had seen earlier. Or it may use all the slots between two sampling instants to send a quantized version of the latest sample. The information utilized by the decoder will be limited by the choice of structure of information transmission adopted at encoder. As the process we consider is Markov in nature, we choose to utilize all the transmission instants between two sampling instants to send information about the latest sample.

3.1. Encoder Structure: Refining the Error Successively

As mentioned earlier, the encoder and decoder that we employ in this work are similar to that in the DPCM scheme [39,51,52]. However, recall that we have multiple transmission opportunities between each sampling instance and the transmissions are delayed. This calls for the need for certain modifications which we explain in the following.

Let $\hat{X}_{ks|t}$ and $\hat{X}_{t|t}$ respectively denote the estimate for X_{ks} and X_t formed at the receiver at time t. Our encoder computes the error in the receiver estimate of the last process sample at each time instant t. Denoting the error at time $t \in I_k$ by $Y_t \triangleq X_{ks} - \hat{X}_{ks|t}$, the encoder quantizes this error Y_t and sends it as communication $C_{ks+1,i}$. At any time instant $t = ks + i$, we suppose that we estimate $\hat{X}_{t|t}$ from $\hat{X}_{ks|t}$. In particular, we suppose that $\hat{X}_{t|t} = \alpha^i \hat{X}_{ks|t}$. Simply speaking, our encoder computes and quantizes the error in the current estimate of the last sample at the decoder, and sends it to the decoder to enable the refinement of the estimate in the next time slot. While we have not been able to establish optimality of this encoder structure, our results will show its asymptotic optimality when the number of dimensions n goes to infinity.

Even within this structural simplification, a very interesting question remains. Since the process is sampled once in s time slots, we have, potentially, nRs bits to encode the latest sample. At any time $t \in \tilde{I}_k$, the receiver has access to $(C_0, \ldots, C_{(k-1)s})$ and the partial codewords $(C_{ks,0}, \ldots, C_{ks,i-1})$ for $i = t - ks$. A simple approach for the encoder is to use the complete codeword to express the latest sample and the decoder can ignore the partial

codewords. This approach will result in slow but very accurate updates of the sample estimates. An alternative fast but loose approach will send nR quantizer codewords to refine estimates in every communication slot. Should we prefer fast but loose estimates or slow but accurate ones? Our results will shed light on this conundrum.

3.2. The Choice of Quantizers

In our description of the encoder structure above, we did not specify a key design choice, namely the choice of the quantizer. We will restrict ourselves to using the same quantizer to quantize the error in each round of communication. The precision of this quantizer will depend on whether we choose a fast but loose paradigm or a slow but accurate one. However, the overall structure will remain the same. Roughly speaking, we allow any gain-shape [53] quantizer which separately sends the quantized value of the gain $\|y\|_2$ and the shape $y/\|y\|_2$ for input y. Formally, we use the following abstraction.

Definition 2 ((θ, ε)-quantizer family). *Fix $0 < M < \infty$. For $0 \leqslant \theta \leqslant 1$ and $0 < \varepsilon$, a quantizer Q with dynamic range M specified by a mapping $Q : \mathbb{R}^n \to \{0,1\}^{nR}$ constitutes an nR bit (θ, ε)-quantizer if for every vector $y \in \mathbb{R}^n$ such that $\|y\|_2^2 \leqslant nM^2$, we have*

$$\mathbb{E}\|y - Q(y)\|_2^2 \leqslant \|y\|_2^2 \theta(R) + n\varepsilon^2.$$

Further, for a mapping $\theta : \mathbb{R}_+ \to [0,1]$, which is a decreasing function of rate R, a family of quantizers $Q = \{Q_R : R > 0\}$ constitutes an (θ, ε)-quantizer family if for every R the quantizer Q_R constitutes an nR bit ($\theta(R), \varepsilon$)-quantizer.

The expectation in the previous definition is taken with respect to the randomness in the quantizer, which is assumed to be shared between the encoder and the decoder for simplicity. For instance, in Lemma 5 we study a random codebook based construction for a (θ, ε)-quantizer and we assume that once a random codebook is picked by the encoder, it is made known to the decoder. The parameter M, termed the *dynamic range* of the quantizer, specifies the domain of the quantizer. When the input y does not satisfy $\|y\|_2 \leqslant \sqrt{n}M$, the quantizer simply declares a failure, which we denote by $\perp$. Our tracking code may use any such (θ, ε)-quantizer family. It is typical in any construction of a gain-shape quantizer to have a finite M and $\varepsilon > 0$. Our analysis for finite n will apply to any such (θ, ε)-quantizer family and, in particular, will bring-out the role of the "bias" ε. However, when establishing our optimality result, we instantiate it using a random spherical code to get the desired performance.

3.3. Description of the Successive Update Scheme

All the conceptual components of our scheme are ready. Note that, we focus on updating the estimates of the latest observed sample X_{ks} at the decoder. Our encoder successively updates the estimate of the latest sample at the decoder by quantizing and sending estimates for errors Y_t.

As discussed earlier, we must decide if we prefer a fast but loose approach or a slow but accurate approach for sending error estimates. To carefully examine this tradeoff, we opt for a more general scheme where the nRs bits available between two samples are divided into $m = s/p$ subfragments of length nRp bits each. We use an nRp bit quantizer to refine error estimates for the latest sample X_{ks} (obtained at time $t = ks$) every p slots and send the resulting quantizer codewords as partial tracking codewords $(C_{ks,jp}, ..., C_{ks,(j+1)p-1})$, $0 \leqslant j \leqslant m - 1$. Specifically, the kth codeword transmission interval I_k is divided into m subfragments $I_{k,j}, 1 \leqslant j \leqslant m$ given by

$$I_{k,j} \triangleq \{ks + jp, \dots, ks + (j+1)p - 1\}, \ 0 \leqslant j \leqslant m - 1,$$

and $(C_{ks,jp}, ..., C_{ks,(j+1)p-1})$ is transmitted in communication slots in $I_{k,j}$.

At time instant $t = ks + jp + 1$ the decoder receives the jth subfragment $(C_{ks,t-ks}, t \in I_{k,j})$ of nRp bits and uses it to refine the estimate of the latest source sample X_{ks}. Note that the fast but loose and the slow but accurate regimes described above correspond to $p = 1$ and $p = s$, respectively. In the middle of the interval $I_{k,j}$, the decoder ignores the partially received quantization code and retains the estimate $\hat{X}_{ks}$ of X_{ks} formed at time $ks + (j-1)p + 1$. It forms an estimate of the current state X_{ks+i} by simply scaling $\hat{X}_{ks}$ by a factor of α^i.

Finally, we impose one more additional simplification on the decoder structure. We simply update the estimate by adding to it the quantized value of the error. Thus, the decoder has a simple linear structure.

We can use any nRp bit quantizer (with an abuse of notation, we will use Q_p instead of Q_{Rp} to denote an nRp bit quantizer) Q_p for the n-dimensional error vector, whereby this scheme can be easily implemented in practice if Q_p can be implemented. For instance, we can use any standard gain-shape quantizer. The performance of most quantizers can be analyzed explicitly to render them a (θ, ε)-quantizer family for an appropriate M and function θ. Later, when analyzing the scheme, we will consider a Q_p coming from a (θ, ε)-quantizer family and present a theoretically sound guideline for choosing p.

Recall that we denote the estimate of X_u formed at the decoder at time $t \geqslant u$ by $\hat{X}_{u|t}$. We start by initializing $\hat{X}_{0|0} = 0$ and then proceed using the encoder and the decoder algorithms outlined above. Note that our quantizer Q_p may declare failure symbol $\perp$, in which case the decoder must still yield a nominal estimate. We will simply declare the estimate as (In analysis, we account for all these events as error. Only the probability of failure will determine the contribution of this part to the MSE since the process is mean-square bounded.) 0 once a failure happens.

We give a formal description of our encoder and decoder algorithms below.

The encoder.

1   Initialize $k = 0$, $j = 0$, $\hat{X}_{0|0} = 0$.
2   At time $t = ks + jp$, use the decoder algorithm (to be described below) to form the estimate $\hat{X}_{ks|t}$ and compute the error

$$Y_{k,j} \triangleq X_{ks} - \hat{X}_{ks|t}, \tag{2}$$

where we use the latest sample X_{ks} available at time $t = ks + jp$.
3   Quantize $Y_{k,j}$ to nRp bit as $Q_p(Y_{k,j})$.
4   If quantize failure occurs and $Q_p(Y_{k,j}) = \perp$, send $\perp$ to the receiver and terminate the encoder.
5   Else, send a binary representation of $Q_p(Y_{k,j})$ as the communication $(C_{ks,0}, ..., C_{ks,p-1})$ to the receiver over the next p communication slots (For simplicity, we do not account for the extra message symbol needed for sending $\perp$.).
6   If $j < m - 1$, increase j by 1; else set $j = 0$ and increase k by 1. Go to Step 2.

The decoder.

1   Initialize $k = 0$, $j = 0$, $\hat{X}_{0|0} = 0$.
2   At time $t = ks + jp$, if encoding failure has not occurred until time t, compute

$$\hat{X}_{ks|ks+jp} = \hat{X}_{ks|ks+(j-1)p} + Q_p(Y_{k,j-1}),$$

and output $\hat{X}_{t|t} = \alpha^{t-ks} \hat{X}_{ks|t}$.
3   Else, if encoding failure has occurred and the $\perp$ symbol is received declare $\hat{X}_{s|t} = 0$ for all subsequent time instants $s \geqslant t$.
4   At time $t = ks + jp + i$, for $i \in [p-1]$, output (We ignore the partial quantizer codewords received as $(C_{ks,jp+1}, C_{ks,jp+2}, \ldots, C_{ks,jp+i-1})$ till time t.) $\hat{X}_{t|t} = \alpha^{t-ks} \hat{X}_{ks|ks+jp}$.
5   If $j < m - 1$, increase j by 1; else set $j = 0$ and increase k by 1. Go to Step 2.

Note that the decoder has a simple structure and the principal component of the encoder is the quantizer. Therefore, the complexity of the proposed scheme is dominated by the complexity of the quantization operation and varies with respect to the quantizer chosen.

4. Main Results

We present results in two categories. First, we provide an explicit formula for the asymptotic maxmin tracking accuracy $\delta^*(R, s, \mathbb{X})$. Next, we present a theoretically-founded guideline for selecting a good p for the successive update scheme with a (θ, ε)-quantizer family. Interestingly, the optimal choice may differ from the asymptotically optimal choice of $p = 1$.

4.1. Characterization of the Maxmin Tracking Accuracy

To describe our result, we define the functions $\delta_0 : \mathbb{R}_+ \to [0,1]$ and $g : \mathbb{R}_+ \to [0,1]$ as

$$\delta_0(R) \triangleq \frac{\alpha^2(1 - 2^{-2R})}{(1 - \alpha^2 2^{-2R})}, \text{ for all } R > 0;$$

$$g(s) \triangleq \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)}, \text{ for all } s > 0. \tag{3}$$

Note that $g(s)$ is a decreasing function of s with $g(1) = 1$. The result below shows that, for an appropriate choice of the quantizer, our successive update scheme with $p = 1$ (the *fast but loose* version) achieves an accuracy of $\delta_0(R)g(s)$ asymptotically, universally for all processes in $\mathbb{X}$.

Theorem 1 (Lower bound for maxmin tracking accuracy: The achievability). *For $R > 0$ and $s \in \mathbb{N}$, the asymptotic maxmin tacking accuracy is bounded below as*

$$\delta^*(R, s, \mathbb{X}) \geqslant \delta_0(R)g(s).$$

Furthermore, this bound can be obtained by a successive update scheme with $p = 1$ and appropriately chosen quantizer Q_p.

We provide a proof in Section 6. Note that while we assume that the per dimension fourth moment of the processes in $\mathbb{X}$ is bounded, the asymptotic result above does not depend on that bound. Interestingly, the performance characterized above is the best possible.

Theorem 2 (Upper bound for maxmin tracking accuracy: The converse). *For $R > 0$ and $s \in \mathbb{N}$, the asymptotic maxmin tacking accuracy is bounded above as*

$$\delta^*(R, s, \mathbb{X}) \leqslant \delta_0(R)g(s).$$

Furthermore, the upper bound is obtained by considering a Gauss–Markov process.

We provide a proof in Section 7. Thus, $\delta^*(R, s, \mathbb{X}) = \delta_0(R)g(s)$ with the *fast but loose* successive update scheme being universally (asymptotically) optimal and the Gauss–Markov process being the most difficult process to track. Clearly, the best possible choice of sampling period is $s = 1$ and the highest possible accuracy at rate R is $\delta_0(R)$, whereby we cannot hope for an accuracy exceeding $\delta_0(R)$.

To provide an alternative view of the result, suppose that we fix the achievable tracking accuracy to be $\delta \leqslant \delta_0(R)$ at rate R. Then, the above result says that $g(s) \geqslant (\delta/\delta_0(R))$ where $g(s)$ is a decreasing function of s. Therefore, this result can be interpreted as saying that we cannot subsample at a frequency less than $1/\lfloor g^{-1}(\delta/\delta_0(R)) \rfloor$ for attaining a tracking accuracy $\delta \leqslant \delta_0(R)$.

4.2. Guidelines for Choosing a Good p

The proof of Theorem 1 entails the analysis of the successive update scheme for $p = 1$. In fact, we can analyze this scheme for any $p \in \mathbb{N}$ and for any (θ, ε)-quantizer family; we term this tracking code the *p-successive update* (*p-SU*) scheme. This analysis can provide a simple guideline for the optimal choice of p depending on the performance of the quantizer.

However, there are some technical caveats. A quantizer family will operate only as long as the input y satisfies $\|y\|_2 \leqslant M$. If a y outside this range is observed, the quantizer will declare $\perp$ and the tracking code encoder, in turn, will declare a failure. We denote by τ the stopping time at which encoder failure occurs for the first time, *i.e.*,

$$\tau \triangleq \min\{ks + jp : Q_p(Y_{k,j}) = \perp, 0 \leqslant k, 0 \leqslant j \leqslant m - 1\}.$$

Further, denote by A_t the event that failure does not occur until time t, *i.e.*,

$$A_t \triangleq \{\tau > t\}.$$

We characterize the performance of a p-SU in terms of the probability of encoder failure in a finite time horizon T.

Theorem 3 (Performance of p-SU). *For fixed $\theta, \varepsilon, \beta \in [0, 1]$, consider the p-SU scheme with an nRp bit (θ, ε)-quantizer Q_p, and denote the corresponding tracking code by (ϕ_p, ψ_p). Suppose that for a time horizon $T \in \mathbb{N}$, the tracking code (ϕ_p, ψ_p) satisfies $\mathbb{P}(\tau \leqslant T) \leqslant \beta$. Then,*

$$\sup_{X \in \mathbb{X}_n} \overline{D}^T(\phi_p, \psi_p, X) \leqslant B_T(\theta, \varepsilon, \beta),$$

where $B_T(\theta, \varepsilon, \beta)$ satisfies

$$\limsup_{T \to \infty} B_T(\theta, \varepsilon, \beta) \leqslant \sigma^2 \left[1 - \frac{g(s)\, \alpha^{2p}}{1 - \alpha^{2p}\,\theta} \left(1 - \frac{\varepsilon^2}{\sigma^2} - \theta \right) \right] + \frac{\kappa \beta g(s)}{(1 - \alpha^{2s})} \left(1 - \alpha^{2(s+p)} \frac{(1 - \theta)}{1 - \alpha^{2p}\theta} \right).$$

We provide the proof of this theorem later in Section 5. We remark that β can be made small by choosing M to be large for a quantizer family. Furthermore, the inequality in the upper bound for the MSE in the previous result (barring the dependence on β) comes from the inequality in the definition of a (θ, ε)-quantizer, rendering it a good proxy for the performance of the quantizer. The interesting regime is that of very small β where the encoder failure does not occur during the time horizon of operation. If we ignore the dependence on β, the accuracy of the p-SU does not depend either on s or on the bound for the fourth moment κ. Motivated by these insights, we define the *accuracy-speed* curve of a quantizer family as follows.

Definition 3 (The accuracy-speed curve). *For $\alpha \in [0, 1]$, σ^2, and $R > 0$, the* accuracy-speed *curve for a (θ, ε)-quantizer family Q is given by*

$$\Gamma_Q(p) = \frac{\alpha^{2p}}{1 - \alpha^{2p}\,\theta(Rp)} \left(1 - \frac{\varepsilon^2}{\sigma^2} - \theta(Rp) \right), \quad p > 0.$$

By Theorem 3, it is easy to see that the accuracy (precisely the upper bound on the accuracy) of a p-SU scheme is better when $\Gamma_Q(p)$ is larger. Thus, a good choice of p for a given quantizer family Q is the one that maximizes $\Gamma_Q(p)$ for $1 \leqslant p \leqslant s$.

We conclude by providing accuracy-speed curves for some illustrative examples. To build some heuristics, note that a uniform quantization of $[-M, M]$ has $\theta(R) = 0$ and $\varepsilon = M2^{-R}$. For a gain-shape quantizer, we express a vector $y = \|y\|_2 y_s$ where the shape vector y_s has $\|y_s\|_2 = 1$. An ideal shape quantizer (which only can be shown to exist asymptotically) using R bits per dimension will satisfy $\mathbb{E}\|\hat{y}_s - y_s\|_2^2 \leqslant 2^{-2R}$, similar to the

scalar uniform quantizer. In one of the examples below, we consider gain-shape quantizers with such an ideal shape quantizer.

Example 1. *We begin by considering an ideal quantizer family with $\theta(R) = 2^{-2R}$ and $\varepsilon = 0$. In our asymptotic analysis, we will show roughly that such a quantizer with very small ε exists. For this ideal case, for $R > 0$, the accuracy-speed curve is given by*

$$\Gamma_Q(p) = \frac{\alpha^{2p} - \alpha^{2p}\,\theta(Rp)}{1 - \alpha^{2p}\,\theta(Rp)} = 1 - \frac{1 - \alpha^{2p}}{1 - \alpha^{2p}2^{-Rp}}.$$

It can be seen that $\Gamma_Q(p)$ is decreasing in p whereby the optimal choice of p that maximized $\Gamma_Q(p)$ over $p \in [s]$ is $p = 1$. Heuristically, this justifies why asymptotically the fast but loose successive update scheme is optimal.

Example 2 (Uniform scalar quantization)**.** *In this example, we consider a coordinate-wise uniform quantizer. Since we seek quantizers for inputs $y \in \mathbb{R}^n$ such that $\|y\|_2 \leqslant M\sqrt{n}$, we can only use uniform quantizer of $[-M\sqrt{n}, M\sqrt{n}]$ for each coordinate. For this quantizer, we have $\theta = 0$ and $\varepsilon^2 = nM^2 2^{-2R}$, whereby the accuracy-speed curve is given by $\Gamma_Q(p) = \alpha^{2p}(1 - nM^2 2^{-2R}/\sigma^2)$. Thus, once again, the optimal choice of p that maximizes accuracy is $p = 1$.*

Example 3 (Gain-shape quantizer)**.** *Consider the quantization of a vector $y = ay_s$ where $a = \|y\|_2$. The vector y is quantized by a gain-shape quantizer which quantizes the norm and shape of the vector separately to give $Q(y) = \hat{a}\hat{y}_s$. We use a uniform quantizer within a fixed range $[0, M\sqrt{n}]$ in order to quantize the norm a to $\hat{a}$, where an ideal shape quantizer is employed in quantizing the shape vector y_s. Namely, we assume $\mathbb{E}\|y_s - \hat{y}_s\|_2^2 \leqslant 2^{-2R}$ and $\|\hat{y}_s\| \leqslant 1$. Suppose that we allot ℓ bits out of the total budget of nR bits for norm quantization and the rest for shape quantization. Then, we see that*

$$\mathbb{E}\|y - Q(y)\|_2^2 \leqslant 2a^2 2^{-2(R-\ell/n)} + nM^2 2^{-2\ell-1},$$

whereby $\theta(R) = 2^{-2(R-\ell/n)+1}$ and $\epsilon^2 = M^2 2^{-2\ell-1}$. Thus, the accuracy-speed curve is given by

$$\Gamma_Q(p) = \frac{\alpha^{2p}}{1 - 2\alpha^{2p}2^{-2(Rp-\ell/n)}} \left(1 - \frac{2M^2 2^{-2\ell-1}}{\sigma^2} - 2^{-2(Rp-\ell/n)+1}\right).$$

Note that the optimal choice of p in this case depends on the choice of M.

We illustrated application of our analysis for idealized quantizers, but it can be used to analyze even very practical quantizers, such as the recently proposed almost optimal quantizer in [54].

5. Analysis of the Successive Update Scheme

From the discussion in Section 3, we observe that the successive update scheme is designed to refine the estimate of X_{ks} in each interval $\tilde{I}_k$. This fact helps us in establishing a recursive relation for $D_t(\phi_p, \psi_p, X)$, $t \in \tilde{I}_k$ in terms of $D_{ks}(\phi_p, \psi_p, X)$ which is provided next.

Lemma 1. *For a time instant $t = ks + jp + i, 0 \leqslant j \leqslant m-1, 0 \leqslant i \leqslant p-1$ and $k \geqslant 0$, let (ϕ_p, ψ_p) denote the tracking code of a p-SU scheme employing an nRp bit (θ, ϵ)-quantizer. Assume that $\mathbb{P}(A_t^c) \leqslant \beta^2$. Then, we have*

$$D_t(\phi_p, \psi_p, X) \leqslant \alpha^{2(t-ks)}\theta^j D_{ks}(\phi_p, \psi_p, X) + \sigma^2(1 - \alpha^{2(t-ks)}) + \frac{\alpha^{2(t-ks)}(1-\theta^j)\varepsilon^2}{(1-\theta)} + \alpha^{2(t-ks)}\kappa\beta.$$

Proof. From the evolution of the AR[1] process defined in (1), we see that $X_t = \alpha^{t-ks} X_{ks} + \sum_{u=ks+1}^{t} \alpha^{t-u} \xi_u$. Further for the *p*-SU scheme, we know that $\hat{X}_{t|t} = \alpha^{t-ks} \hat{X}_{ks|ks+jp}$ at each instant $t = ks + jp + i$. Therefore, we have

$$X_t - \hat{X}_{t|t} = \alpha^{t-ks}(X_{ks} - \hat{X}_{ks|ks+jp}) + \sum_{u=ks+1}^{t} \alpha^{t-u} \xi_u.$$

Since the estimate $\hat{X}_{ks|ks+jp}$ is a function of samples $(X_0, \ldots, X_{ks})$, and the sequence $(\xi_u, u \geqslant ks)$ is independent of the past, we obtain the per dimension MSE as

$$D_t(\phi_p, \psi_p, X) = \frac{\alpha^{2(t-ks)}}{n} \mathbb{E} \|X_{ks} - \hat{X}_{ks|ks+jp}\|_2^2 + \sigma^2(1 - \alpha^{2(t-ks)}).$$

Further, we divide the error into two terms based on occurrence of the failure event as follows:

$$D_t(\phi_p, \psi_p, X) = \frac{\alpha^{2(t-ks)}}{n} \left[\mathbb{E}[\|X_{ks} - \hat{X}_{ks|ks+jp}\|_2^2 \mathbb{1}_{A_t}] + \mathbb{E}[\|X_{ks} - \hat{X}_{ks|ks+jp}\|_2^2 \mathbb{1}_{A_t^c}] \right] \quad (4)$$
$$+ \sigma^2(1 - \alpha^{2(t-ks)}).$$

Recall that at each instant $t = ks + jp$, we refine the estimate $\hat{X}_{ks|ks+(j-1)p}$ of X_{ks} to $\hat{X}_{ks|ks+jp} = (\hat{X}_{ks|ks+(j-1)p} + Q_p(Y_{k,j-1})) \mathbb{1}_{A_t}$. Upon substituting this expression for $\hat{X}_{ks|ks+jp}$, we obtain

$$\mathbb{E}\left[\|X_{ks} - \hat{X}_{ks|ks+jp}\|_2^2 \mathbb{1}_{A_t}\right] = \mathbb{E}[\|Y_{k,j-1} - Q_p(Y_{k,j-1})\|^2 \mathbb{1}_{A_t}]$$

$$\leqslant \theta \mathbb{E}[\|X_{ks} - \hat{X}_{ks|ks+(j-1)p}\|_2^2 \mathbb{1}_{A_t}] + n\varepsilon^2,$$

where the identity uses the definition of error $Y_{k,j-1}$ given in (2) and the inequality holds since Q_p is a (θ, ε)-quantizer. Repeating the previous step recursively, we get

$$\frac{1}{n} \mathbb{E}[\|X_{ks} - \hat{X}_{ks|ks+jp}\|_2^2 \mathbb{1}_{A_t}] \leqslant \theta^j \cdot \frac{1}{n} \mathbb{E}[\|X_{ks} - \hat{X}_{ks|ks}\|_2^2 \mathbb{1}_{A_t}] + \frac{1 - \theta^j}{1 - \theta} \cdot \varepsilon^2$$

$$\leqslant \theta^j \cdot \frac{1}{n} \mathbb{E}\|X_{ks} - \hat{X}_{ks|ks}\|_2^2 + \frac{1 - \theta^j}{1 - \theta} \cdot \varepsilon^2,$$

which is the same as

$$\frac{1}{n} \mathbb{E}[\|X_{ks} - \hat{X}_{ks|ks+jp}\|_2^2 \mathbb{1}_{A_t}] \leqslant \theta^j \cdot D_{ks}(\phi_p, \psi_p, X) + \frac{1 - \theta^j}{1 - \theta} \cdot \varepsilon^2.$$

Moving to the error term $\mathbb{E}[\|X_{ks} - \hat{X}_{ks|ks+jp}\|_2^2 \mathbb{1}_{A_t^c}]$ when encoder failure occurs, recall that the decoder sets the estimate to 0 in the event of an encoder failure. Thus, using the Cauchy–Schwarz inequality, we get

$$\frac{1}{n} \mathbb{E}[\|X_{ks} - \hat{X}_{ks|ks+jp}\|_2^2 \mathbb{1}_{A_t^c}] = \frac{1}{n} \mathbb{E}[\|X_{ks}\|^2 \mathbb{1}_{A_t^c}]$$

$$\leqslant \frac{1}{n} \sqrt{\mathbb{E}[\|X_{ks}\|^4] P(A_t^c)}$$

$$\leqslant \kappa \beta.$$

Substituting the two bounds above in (4), we get the result. □

The following recursive bound can be obtained using almost the same proof as that of Lemma 1; we omit the details.

Lemma 2. *Let (ϕ_p, ψ_p) denote the tracking code of a p-SU scheme employing an nRp bit (θ, ϵ)-quantizer. Assume that $\mathbb{P}(A_i^c) \leqslant \beta^2$. Then, we have*

$$D_{ks}(\phi_p, \psi_p, \mathbb{X}_n) \leqslant \alpha^{2s}\theta^m D_{(k-1)s}(\phi_p, \psi_p, \mathbb{X}_n) + \sigma^2(1 - \alpha^{2s}) + \alpha^{2s}\varepsilon^2 \frac{(1 - \theta^m)}{(1 - \theta)} + \alpha^{2s}\kappa\beta.$$

We also need the following technical observation.

Lemma 3. *For a sequence $(X_k \in \mathbb{R} : k \in \mathbb{Z}_+)$ that satisfies sequence of upper bounds*

$$X_k \leqslant aX_{k-1} + b, \qquad \forall k \in \mathbb{Z}_+,$$

with constants $a, b \in \mathbb{R}$ such that b is finite and $a \in (-1, 1)$, we have

$$\lim_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} X_k \leqslant \frac{b}{1 - a}.$$

Proof. From the sequence of upper bounds, we can inductively show that

$$X_k \leqslant a^k X_0 + b\frac{1 - a^k}{1 - a}, \qquad \forall k \in \mathbb{Z}_+.$$

Averaging X_k over the horizon $\{0, \ldots, K-1\}$, we get

$$\frac{1}{K} \sum_{k=0}^{K-1} X_k \leqslant \frac{1 - a^K}{K(1 - a)}\left(X_0 - \frac{b}{1 - a}\right) + \frac{b}{1 - a}.$$

From the finiteness of X_0, b and the fact that $|a| < 1$, the result follows by taking the limit K growing arbitrarily large on both sides. □

We are now in a position to prove Theorem 3.

Proof of Theorem 3. We begin by noting that, without any loss of generality, we can restrict to $T = Ks$. This holds since the contributions of the error term within the fixed interval I_K are bounded. For $T = Ks$, the time duration $\{0, \ldots, T\}$ can be partitioned into intervals $(I_k, k + 1 \in [K])$. Therefore, we can write the average MSE per dimension for the p-SU scheme for time-horizon $T = Ks$ as

$$\overline{D}_T(\phi_p, \psi_p, X) = \frac{1}{Ks} \sum_{k=0}^{K-1} \sum_{j=0}^{m-1} \sum_{i=0}^{p-1} D_{ks+jp+i}(\phi_p, \psi_p, X).$$

From the upper bound for per dimension MSE given in Lemma 1, we get

$$\sum_{i=0}^{p-1} D_{ks+jp+i}(\phi_p, \psi_p, X) \leqslant \sum_{i=0}^{p-1} \left[\alpha^{2(jp+i)}\theta^j D_{ks}(\phi_p, \psi_p, X) + \sigma^2(1 - \alpha^{2(jp+i)}) + \frac{\alpha^{2(jp+i)}(1 - \theta^j)\varepsilon^2}{(1 - \theta)} + \alpha^{2(jp+i)}\kappa\beta\right]$$

$$= \alpha^{2jp}\frac{1 - \alpha^{2p}}{1 - \alpha^2}\left(\theta^j D_{ks}(\phi_p, \psi_p, X) + \frac{(1 - \theta^j)\varepsilon^2}{(1 - \theta)} + \kappa\beta - \sigma^2\right) + p\sigma^2.$$

Summing the expression above over $j \in \{0, \ldots, m-1\}$ and $k \in \{0, \ldots, K-1\}$, and dividing by T, we get

$$\overline{D}_T(\phi_p, \psi_p, X) \leqslant \sigma^2 + \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)}\left(\kappa\beta - \sigma^2 + \frac{\varepsilon^2}{1 - \theta}\right)$$

$$+ \frac{(1 - \alpha^{2p})}{s(1 - \alpha^2)} \frac{(1 - \alpha^{2s}\theta^m)}{(1 - \alpha^{2p}\theta)} \Big(\frac{1}{K} \sum_{k=0}^{K-1} D_{ks}(\phi_p, \psi_p, X) - \frac{\varepsilon^2}{1 - \theta} \Big).$$

It follows by Lemma 2 that

$$\overline{D}_T(\phi_p, \psi_p, X) \leqslant \sigma^2 + \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)} \Big(\kappa\beta - \sigma^2 + \frac{\varepsilon^2}{1 - \theta} \Big) \tag{5}$$

$$+ \frac{(1 - \alpha^{2p})}{s(1 - \alpha^2)} \frac{(1 - \alpha^{2s}\theta^m)}{(1 - \alpha^{2p}\theta)} \Big(\sup_{\{a_k\}_{k \geqslant 0} \in \mathcal{A}} \frac{1}{K} \sum_{k=0}^{K-1} a_k - \frac{\varepsilon^2}{1 - \theta} \Big),$$

where the $\mathcal{A}$ denotes the set of $\{a_k\}_{k \geqslant 0}$ satisfying

$$a_k \leqslant \alpha^{2s}\theta^m a_{k-1} + \sigma^2(1 - \alpha^{2s}) + \alpha^{2s}\kappa\beta + \alpha^{2s}\varepsilon^2 \frac{(1 - \theta^m)}{(1 - \theta)}.$$

We denote the right-side of (5) by $B_T(\theta, \epsilon, \beta)$. Noting that by Lemma 3 any sequence $\{a_k\}_{k \geqslant 0} \in \mathcal{A}$ satisfies

$$\lim_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} a_k \leqslant \frac{1}{1 - \alpha^{2s}\theta^m} \Big(\sigma^2(1 - \alpha^{2s}) + \alpha^{2s}\kappa\beta + \alpha^{2s}\varepsilon^2 \frac{(1 - \theta^m)}{(1 - \theta)} \Big),$$

we get that

$$\limsup_{T \to \infty} B_T(\theta, \epsilon, \beta) \leqslant \sigma^2 \Big(1 - \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)} \cdot \frac{\alpha^{2p}(1 - \theta)}{(1 - \alpha^{2p}\theta)} \Big) + \varepsilon^2 \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)} \cdot \frac{\alpha^{2p}}{(1 - \alpha^{2p}\theta)}$$

$$+ \kappa\beta \frac{1}{s(1 - \alpha^2)} \Big(1 - \frac{\alpha^{2(s+p)}(1 - \theta)}{(1 - \alpha^{2p}\theta)} \Big),$$

which completes the proof. $\square$

6. Asymptotic Achievability Using Random Quantizer

With Theorem 3 at our disposal, the proof of achievability can be completed by fixing $p = 1$ and showing the existence of appropriate quantizer. However, we need to handle the failure event, and we address this first. The next result shows that the failure probability depends on the quantizer only through M.

Lemma 4. *For fixed T and n, consider the p-SU scheme with $p = 1$ and an nR bit (θ, ϵ)-quantizer Q with dynamic range M. Then, for every $\eta > 0$, there exists an M_0 independent of n such that for all $M \geqslant M_0$, we get*
$$\mathbb{P}(A_T^c) \leqslant \eta.$$

Proof. The event A_T (of encoder failure not happening until time T for the successive update scheme) occurs when the errors $Y_{k,j}$ satisfies $\|Y_{k,j}\|_2^2 \leqslant nM^2$, for every $k \geqslant 0$ and $0 \leqslant j \leqslant s - 1$ such that $t = ks + j \leqslant T$. For brevity, we denote by Y_t the error random variable $Y_{k,j}$ and $Y^{t-1} = (Y_1, ..., Y_{t-1})$. We note that

$$\mathbb{P}(A_T^c) = \mathbb{P}(A_{T-1}^c) + \mathbb{P}(A_{T-1} \cap A_T^c)$$

$$= \mathbb{P}(A_{T-1}^c) + \mathbb{P}\Big(A_{T-1} \cap \{\|Y_T\|_2^2 > nM^2\} \Big)$$

$$= \mathbb{P}(A_{T-1}^c) + \mathbb{E}\big[\mathbb{1}_{A_{T-1}} \mathbb{P}\big(\|Y_T\|_2^2 > nM^2 | Y^{T-1} \big) \big]$$

$$\leqslant \mathbb{P}(A_{T-1}^c) + \mathbb{E}\Big[\mathbb{1}_{A_{T-1}} \frac{\mathbb{E}[\|Y_T\|_2^2 | Y^{T-1}]}{nM^2} \Big]$$

$$= \mathbb{P}(A^c_{T-1}) + \frac{\mathbb{E}[\|Y_T\|_2^2 \mathbb{1}_{A_{T-1}}]}{nM^2}.$$

Note that the inequality follows from Markovs inequality. Further we have $Y_T = Y_{T-1} - Q(Y_{T-1})$ under A_{T-1}, whereby

$$\mathbb{E}[\|Y_T\|_2^2 \mathbb{1}_{A_{T-1}}] \leqslant \theta \cdot \mathbb{E}[\|Y_{T-1}\|_2^2 \mathbb{1}_{A_{T-1}}] + n\varepsilon^2.$$

Denoting by β_T^2 the probability $\mathbb{P}(A^c_T)$, the previous two inequalities imply

$$\beta_T^2 \leqslant \beta_{T-1}^2 + \frac{\theta}{nM^2}\mathbb{E}[\|Y_{T-1}\|_2^2] + \frac{\varepsilon^2}{M^2}.$$

We saw earlier in the proof of Lemma 1 that $\mathbb{E}[\|Y_{T-1}\|_2^2]/n$ depends only on the probability β_{T-1}^2 that failure does not occur until time $T - 1$. Proceeding as in that proof, we get

$$\beta_T^2 \leqslant \beta_{T-1}^2 + \frac{1}{M^2}(c_1\beta_{T-1} + c_2),$$

where c_1 and c_2 do not depend on n. Therefore, there exists M_0 independent of n such that for all M exceeding M_0 we have

$$\beta_T^2 \leqslant \beta_{T-1}^2 + \eta,$$

which completes the proof by summing over T. □

The bound above is rather loose, but it suffices for our purpose. In particular, it says that we can choose M sufficiently large to make probability of failure until time T less than any β^2, whereby Theorem 3 can be applied by designing a quantizer for this M. Indeed, we can use the quantizer of unit sphere from [1,2], along with a uniform quantizer for gain (which lies in $[-M, M]$) to get the following performance. In fact, we will show that a deterministic quantizer with the desired performance exists. Note that we already considered such a quantizer in Example 3. However, the analysis there was slightly loose and it assumed the existence of an ideal shape quantizer.

Lemma 5. *For every $R, \varepsilon, \gamma, M > 0$, there exists an nR bit $(2^{-2(R-\gamma)}, \varepsilon)$-quantizer with dynamic range M, for all n sufficiently large.*

Proof. We first borrow a classic construction from [1,2], which gives us our desired shape quantizer. Denote by $\mathbb{S}_n$ the $(n-1)$-dimensional unit sphere $\{y \in \mathbb{R}^n : \|y\|_2 = 1\}$. For every $\gamma > 0$ and n sufficiently large, it was shown in [1,2] that there exist 2^{nR} vectors $\mathcal{C}$ in $\mathbb{S}_n$ such that for every $y \in \mathbb{S}_n$ we can find $y' \in \mathcal{C}$ satisfying

$$\langle y, y' \rangle \geqslant \sqrt{1 - 2^{-2(R-\gamma)}}.$$

Denoting $\cos\theta = \sqrt{1 - 2^{-2(R-\gamma)}}$, consider the shape quantizer $Q_R(y)$ from [2] given by

$$Q_R(y) \triangleq \cos\theta \cdot \arg\min_{y' \in \mathcal{C}} \|y - y'\|_2^2$$
$$= \cos\theta \cdot \arg\max_{y' \in \mathcal{C}} \langle y, y' \rangle, \quad \forall y \in \mathbb{S}_n.$$

Note that we shrink the length of y' by a factor of $\cos\theta$, which will be seen to yield the gain over the analysis in Example 3.

We append to this shape quantizer the uniform gain quantizer $q_M : [0, M] \to [0, M]$, which quantizes the interval $[0, M]$ uniformly into subintervals of length ε. Specifically, $q_M(a) = \varepsilon\lfloor a/\varepsilon \rfloor$ and the corresponding index is given by $\lfloor a/\varepsilon \rfloor$. We represent this index using its $\ell \triangleq \lceil \log(M/\varepsilon) \rceil$ bit binary representation and denote this mapping by $\phi^M : [0, M] \to \{0, 1\}^\ell$.

For every $y \in \mathbb{R}^n$ such that $\|y\|_2^2 \leqslant nM^2$, we consider the quantizer

$$Q(y) = \sqrt{n} \cdot q_M \left(\frac{\|y\|_2}{\sqrt{n}} \right) \cdot Q_R \left(\frac{y}{\|y\|_2} \right).$$

For this quantizer, for every $y \in \mathbb{R}^n$ with $\|y\|_2^2 = nB^2$ such that $B \leqslant M$, we have

$$
\begin{aligned}
\|y - Q(y)\|_2^2 &= \|y\|_2^2 + \|Q(y)\|_2^2 - 2\langle y, Q(y) \rangle \\
&= nB^2 + n\hat{B}^2 \cos^2 \theta - 2nB\hat{B} \cos \theta \langle \tilde{y}, Q_R(\tilde{y}) \rangle \\
&\leqslant nB^2 + n\hat{B}^2 \cos^2 \theta - 2nB\hat{B} \cos^2 \theta \\
&= nB^2 \sin^2 \theta + n(B - \hat{B})^2 \cos^2 \theta \\
&\leqslant nB^2 \sin^2 \theta + n\varepsilon^2 \cos^2 \theta \\
&\leqslant nB^2 2^{-2(R-\gamma)} + n\varepsilon^2,
\end{aligned}
$$

where the first inequality uses the covering property of $\mathcal{C}$. Therefore, Q constitutes an $nR + \ell$ bit $2^{-2(R-\gamma),\varepsilon}$-quantizer with dynamic range M, for all n sufficiently large. Note that this quantizer is a deterministic one. $\square$

Proof of Theorem 1. For any fixed β and ε, we can make the probability of failure until time T less than β by choosing M sufficiently large. Further, for any fixed $R, \gamma > 0$, by Lemma 5, we can choose n sufficiently large to get an nR bit $(2^{-2(R-\gamma)}, \varepsilon)$-quantizer for vectors y with $\|y\|_2^2 \leqslant nM^2$. Therefore, by Theorem 3 applied for $p = 1$, we get that

$$\delta^*(R, s, \mathbb{X}) \geqslant \frac{g(s)\,\alpha^2}{1 - \alpha^2 2^{-2(R-\gamma)}} \left(1 - \frac{\varepsilon^2}{\sigma^2} - 2^{-2(R-\gamma)} \right) - \frac{\kappa\beta g(s)}{\sigma^2(1 - \alpha^{2s})} \left(1 - \alpha^{2(s+1)} \frac{1 - \theta}{1 - \alpha^2 \theta} \right).$$

The proof is completed upon taking the limits as ε, γ, and β go to 0. $\square$

7. Converse Bound: Proof of Theorem 2

The proof is similar to the converse proof in [55], but now we need to handle the delay per transmission. We rely on the properties of *entropy power* of a random variable. Recall that for a continuous random variable X taking values in $\mathbb{R}^n$, the entropy power of X is given by

$$\mathcal{N}(X) = \frac{1}{2\pi e} \cdot 2^{(2/n)\,h(X)},$$

where $h(X)$ is the differential entropy of X.

Consider a tracking code (ϕ, ψ) of rate R and sampling period s and a process $X \in \mathbb{X}_n$. We begin by noting that the state at time t is related to the state at time $t + i$ as

$$X_{t+i} = \alpha^i X_t + \sum_{j=0}^{i-1} \alpha^j \xi_{t+i-j},$$

where the noise $\sum_{j=0}^{i-1} \alpha^j \xi_{t+i-j}$ is independent of X_t (and the past states). In particular, for $t = ks + i, 1 \leqslant i < s$, we get

$$
\begin{aligned}
\mathbb{E}\left[\|X_t - \hat{X}_{t|t}\|_2^2 \right] &= \mathbb{E}\left[\|\alpha^i X_{ks} - \hat{X}_{t|t}\|_2^2 \right] + \sum_{j=0}^{i-1} \alpha^{2j} \mathbb{E}\left[\|\xi_{ks+i-j}\|_2^2 \right] \\
&= \alpha^{2i} \mathbb{E}\left[\|X_{ks} - \tilde{X}_t\|_2^2 \right] + n(1 - \alpha^{2i})\sigma^2.
\end{aligned}
$$

where we define $\tilde{X}_t := \alpha^{-i}\hat{X}_{t|t}$ and the first identity uses the orthogonality of noise added in each round from the previous states and noise. The second equality follows directly by substituting the variance of the components of the process ξ_{ks+i-j} and simplifying

the summation. Since the Gaussian distribution has the maximum differential entropy among all continuous random variables with a given variance, and the entropy power for a Gaussian random variable equals its variance, we get that

$$\sigma^2(1 - \alpha^2) \geqslant \mathcal{N}(\xi_{t+i}).$$

Therefore, the previous bound for tracking error yields

$$D_{ks+i}(\phi, \psi, X) \geqslant \alpha^{2i} \frac{1}{n} \mathbb{E}\left[\|X_{ks} - \tilde{X}_{ks+i}\|_2^2\right] + \frac{(1 - \alpha^{2i})}{(1 - \alpha^2)} \mathcal{N}(\xi_{ks+i})$$

$$= \alpha^{2i} \frac{1}{n} \mathbb{E}\left[\|X_{ks} - \tilde{X}_{ks+i}\|_2^2\right] + \frac{(1 - \alpha^{2i})}{(1 - \alpha^2)} \mathcal{N}(\xi_1), \tag{6}$$

where the identity uses the assumption that ξ_t are identically distributed for all t. Taking average of these terms for $t = 0, .., T$, we get

$$\overline{D}_T(\phi, \psi, X) = \frac{1}{nKs} \sum_{k=0}^{K-1} \sum_{i=ks}^{(k+1)s-1} \mathbb{E}\left[\|X_i - \hat{X}_{i|i}\|_2^2\right]$$

$$\geqslant \frac{1}{nKs} \sum_{k=0}^{K-1} \sum_{i=0}^{s-1} \alpha^{2i} \mathbb{E}\left[\|X_{ks} - \tilde{X}_{ks+i}\|_2^2\right] + \frac{\mathcal{N}(\xi_1)}{(1 - \alpha^2)} \left(1 - \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)}\right).$$

Note that $\tilde{X}_{ks+i}$s act as estimates of X_{ks} which depend on the communication received by the decoder until time $ks + i$. We denote the communication received at time t by C_{t-1}, whereby $\tilde{X}_{ks+i}$ depends only on $C_1, ..., C_{ks+i-1}$. In particular, the communication $C_{ks}, ..., C_{ks+i-1}$ was sent as a function of X_{ks}, the sample seen at time $t = ks$.

From here on, we proceed by invoking the "entropy power bounds" for the MSE terms. For random variables X and Y such that $P_{X|Y}$ has a conditional density, the conditional entropy power is given by $\mathcal{N}(X|Y) = 1/(2\pi e)2^{2h(X|Y)/n}$. (The conditional differential entropy $h(X|Y)$ is given by $\mathbb{E}\left[h(P_{X|Y})\right]$.) Bounding MSE terms by entropy power is a standard step that allows us to track reduction in error due to a fixed amount of communication.

We begin by using the following standard bound (see [56], Chapter 10): (It follows simply by noting that Gaussian maximizes differential entropy among all random variables with a given second moment and that $h(X) - h(X|Y) \leqslant H(Y) = nR$.) For a continuous random variable X and a discrete random variable Y taking $\{0, 1\}^{nR}$ values, let $\hat{X}$ be any function of Y. Then, it holds that

$$\frac{1}{n} \mathbb{E}\|X - \hat{X}\|^2 \geqslant 2^{-2R} \mathcal{N}(X). \tag{7}$$

We apply this result to X_{ks} given C^{ks-1} in the role of X and the communication $C_{ks}, .., C_{ks+i-1}$ in the role of Y. The previous bound and Jensen's inequality yield

$$\frac{1}{n} \mathbb{E}\left[\|X_{ks} - \tilde{X}_{ks+i}\|_2^2\right] \geqslant 2^{-2Ri} \mathbb{E}[\mathcal{N}(X_{ks}|C^{ks-1})].$$

Next, we recall the entropy power inequality (cf. [56]): For independent X_1 and X_2, $\mathcal{N}(X_1 + X_2) \geqslant \mathcal{N}(X_1) + \mathcal{N}(X_2)$. Noting that $X_{ks} = \alpha^s X_{(k-1)s} + \sum_{j=0}^{s-1} \alpha^j \xi_{ks-j}$, where $\{\xi_i\}$ is an *iid* zero-mean random variable independent of $X_{(k-1)s}$, and that C^{ks-1} is a function of $X_1, ..., X_{(k-1)s}$, we get

$$\mathcal{N}(X_{ks}|C^{ks-1}) \geqslant \mathcal{N}(\alpha^s X_{(k-1)s}|C^{ks-1}) + \mathcal{N}(\xi_{ks}) \frac{(1 - \alpha^{2s})}{(1 - \alpha^2)}$$

$$= \alpha^{2s} \mathcal{N}(X_{(k-1)s}|C^{ks-1}) + \mathcal{N}(\xi_1) \frac{(1 - \alpha^{2s})}{(1 - \alpha^2)},$$

where the previous identities utilizes the scaling property of differential entropy. Upon combining the bounds given above and simplifying, we get

$$\overline{D}_T(\phi, \psi, X) \geqslant \frac{\alpha^{2s}(1 - \alpha^{2s}2^{-2Rs})}{s(1 - \alpha^2 2^{-2R})} \cdot \frac{1}{K} \sum_{k=0}^{K-1} \mathbb{E}[\mathcal{N}(X_{(k-1)s}|C^{ks-1})]$$
$$+ \frac{\mathcal{N}(\xi_1)}{(1 - \alpha^2)} \left(1 + \frac{(1 - \alpha^{2s})(1 - \alpha^{2s}2^{-2Rs})}{s(1 - \alpha^2 2^{-2R})} - \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)}\right). \quad (8)$$

Finally, note that the terms $\mathcal{N}(X_{(k-1)s}|C^{ks-1})$ are exactly the same as that considered in [55] (eqn. 11e) since they correspond to recovering $X_{(k-1)s}$ using communication that can depend on it. Therefore, a similar expression holds here, for the sampled process $\{X_{ks} : k \in \mathbb{N}\}$. Using the recursive bound for the tracking error in (6) and (7), we adapt the results of [55] (eqn. 11) for our case to obtain

$$\mathbb{E}[\mathcal{N}(X_{(k-1)s}|C^{ks-1})] \geqslant d^*_{k-1},$$

where the quantity d^*_k is given by the recursion

$$d^*_k = 2^{-2Rs}\left(\alpha^{2s}d^*_{k-1} + \mathcal{N}(\xi_1)\frac{(1 - \alpha^{2s})}{(1 - \alpha^2)}\right),$$

with $d^*_0 = 0$.

The bound obtain above holds for any given process $X \in \mathbb{X}_n$. To obtain the best possible bound we substitute ξ_1 to be a Gaussian random variable, since that would maximize $\mathcal{N}(\xi_1)$. Specifically, we set $\{\xi_k\}$ to be a Gaussian random variable with zero mean and variance σ^2 to get $\mathcal{N}(\xi) = \sigma^2(1 - \alpha^2)$. Thus, taking supremum over all distributions on both sides of (8), we have

$$\sup_{X \in \mathbb{X}_n} \overline{D}_T(\phi, \psi, X) \geqslant$$
$$\frac{\alpha^{2s}(1 - \alpha^{2s}2^{-2Rs})}{s(1 - \alpha^2 2^{-2R})} \cdot \frac{1}{K} \sum_{k=0}^{K-1} d^*_{k-1} + \sigma^2\left(1 + \frac{(1 - \alpha^{2s})(1 - \alpha^{2s}2^{-2Rs})}{s(1 - \alpha^2 2^{-2R})} - \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)}\right),$$

where

$$d^*_k = 2^{-2Rs}\left(\alpha^{2s}d^*_{k-1} + \sigma^2(1 - \alpha^{2s})\right),$$

with $d^*_0 = 0$. For this sequence d^*_k, we can see that (cf. [55] (Corollary 1))

$$\limsup_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} d^*_{k-1} = \lim_{K \to \infty} d^*_k = \frac{\sigma^2(1 - \alpha^{2s})2^{-2Rs}}{(1 - \alpha^{2s}2^{-2Rs})}.$$

Therefore, we have obtained

$$\limsup_{T \to \infty} \sup_{X \in \mathbb{X}_n} \overline{D}_T(\phi, \psi, X) \geqslant \sigma^2\left(\frac{(1 - \alpha^{2s})\alpha^{2s}2^{-2Rs}}{s(1 - \alpha^2 2^{-2R})} + 1 - \frac{(1 - \alpha^{2s})}{s(1 - \alpha^2)} + \frac{(1 - \alpha^{2s})(1 - \alpha^{2s}2^{-2Rs})}{s(1 - \alpha^2 2^{-2R})}\right)$$
$$= \sigma^2\left(1 - g(s)\delta_0(R)\right).$$

As the bound obtained above holds for all tracking codes (ϕ, ψ), it follows that $\delta^*(R, s, \mathbb{X}) \leqslant g(s)\delta_0(R)$.

8. Discussion

We restricted our treatment to an AR[1] process with uncorrelated components. This restriction is for clarity of presentation, and some of the results can be extended to AR[1] processes with correlated components. In this case, the decoder will be replaced by a

Kalman-like filter in the manner of [35]. A natural extension of this work is the study of an optimum transmission strategy for an AR[*n*] process in the given setting. In an AR[*n*] process, the strategy of refining the latest sample is clearly not sufficient as the value of the process at any time instant is dependent on the past *n* samples. If the sampling is periodic, even the encoder does not have access to all these *n* samples unless we take a sample at every instant. A viable alternative is to take *n* consecutive samples at every sampling instant. However, even with this structure on the sampling policy, it is not clear how the information must be transmitted. A systematic analysis of this problem is an interesting area of future research.

Another setting which is not discussed in the current work is where the transmissions are of nonuniform rates. Throughout our work, we have assumed periodic sampling and transmissions at a fixed rate. For the scheme presented in this paper, it is easy to see from our analysis that only the total number of bits transmitted in each sampling interval matters, when the dimension is sufficiently large. That is, for our scheme, even framing each packet (sent in each communication slot) using unequal number of bits will give the same performance as that for equal packet size, if the overall bit-budget per sampling period is fixed. A similar phenomenon was observed in [39], which allowed the extension of some of their analysis to erasure channels with feedback. We remark that a similar extension is possible for some of our results, too. This behavior stems from the use of successive batches of bits to successively refine the estimate of a single sample within any sampling interval, whereby at the end of the sampling interval the error corresponds to roughly that for a quantizer using the total number of bits sent during the interval. In general, a study of nonuniform rates for describing each sample, while keeping bits per time-slot fixed, will require us to move beyond uniform sampling. This, too, is an interesting research direction to pursue.

Finally, we remark that the encoder structure we have imposed, wherein the error in the estimate of the latest sample is refined at each instant, is optimal only asymptotically and is justified only heuristically for fixed dimensions. Even for one dimensional observation it is not clear if this structure is optimal. We believe that this is a question of fundamental interest which remains open.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR[1]   Auto-Regressive process of order 1

References

1. Wyner, A.D. Random packings and coverings of the unit n-sphere. *Bell Syst. Tech. J.* **1967**, *46*, 2111–2118. [CrossRef]
2. Lapidoth, A. On the role of mismatch in rate distortion theory. *IEEE Trans. Inf. Theory* **1997**, *43*, 38–47. [CrossRef]
3. Witsenhausen, H.S. On the Structure of Real-Time Source Coders. *Bell Syst. Tech. J.* **1979**, *58*, 1437–1451. [CrossRef]
4. Teneketzis, D. On the Structure of Optimal Real-Time Encoders and Decoders in Noisy Communication. *IEEE Trans. Inf. Theory* **2006**, *52*, 4017–4035. [CrossRef]
5. Mahajan, A.; Teneketzis, D. On Real-Time Communication Systems with Noisy Feedback. In Proceedings of the 2007 IEEE Information Theory Workshop, Tahoe City, CA, USA, 2–6 September 2007; pp. 283–288.
6. Walrand, J.C.; Varaiya, P. Optimal causal coding-decoding problems. *IEEE Trans. Inf. Theory* **1983**, *29*, 814–819. [CrossRef]
7. Yuksel, S. On optimal causal coding of partially observed Markov sources in single and multiterminal settings. *IEEE Trans. Inf. Theory* **2012**, *59*, 424–437. [CrossRef]
8. Linder, T.; Yüksel, S. On optimal zero-delay coding of vector Markov sources. *IEEE Trans. Inf. Theory* **2014**, *60*, 5975–5991. [CrossRef]
9. Wood, R.G.; Linder, T.; Yüksel, S. Optimal zero delay coding of Markov sources: Stationary and finite memory codes. *IEEE Trans. Inf. Theory* **2017**, *63*, 5968–5980. [CrossRef]
10. Dobrushin, R.; Tsybakov, B. Information transmission with additional noise. *IRE Trans. Inf. Theory* **1962**, *8*, 293–304. [CrossRef]
11. Wolf, J.; Ziv, J. Transmission of noisy information to a noisy receiver with minimum distortion. *IEEE Trans. Inf. Theory* **1970**, *16*, 406–411. [CrossRef]
12. Witsenhausen, H. Indirect rate distortion problems. *IEEE Trans. Inf. Theory* **1980**, *26*, 518–521. [CrossRef]
13. Mohammadi, E.; Fallah, A.; Marvasti, F. Sampling and Distortion Tradeoffs for Indirect Source Retrieval. *IEEE Trans. Inf. Theory* **2017**, *63*, 6833–6848. [CrossRef]
14. Zamir, R.; Feder, M. Rate-distortion performance in coding bandlimited sources by sampling and dithered quantization. *IEEE Trans. Inf. Theory* **1995**, *41*, 141–154. [CrossRef]
15. Mashiach, A.; Zamir, R. Entropy-coded quantization of periodic nonuniform samples. In Proceedings of the 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, 14–17 November 2012; pp. 1–5.
16. Mashiach, A.; Zamir, R. Noise-shaped quantization for nonuniform sampling. In Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, 7–12 July 2013; pp. 1187–1191.
17. Kipnis, A.; Goldsmith, A.J.; Eldar, Y.C.; Weissman, T. Distortion rate function of sub-Nyquist sampled Gaussian sources. *IEEE Trans. Inf. Theory* **2015**, *62*, 401–429. [CrossRef]
18. Wong, W.S.; Brockett, R.W. Systems with finite communication bandwidth constraints. I. State estimation problems. *IEEE Trans. Autom. Control* **1997**, *42*, 1294–1299. [CrossRef]
19. Nair, G.N.; Evans, R.J. State estimation via a capacity-limited communication channel. In Proceedings of the 36th IEEE Conference on Decision and Control, San Diego, CA, USA, 12 December 1997; Volume 1, pp. 866–871.
20. Nair, G.N.; Evans, R.J. State estimation under bit-rate constraints. In Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No.98CH36171), Tampa, FL, USA, 18 December 1998; Volume 1, pp. 251–256.
21. Dokuchaev, N.G.; Savkin, A.V. Recursive state estimation via limited capacity communication channels. In Proceedings of the 38th IEEE Conference on Decision and Control, Phoenix, AZ, USA, 7–10 December 1999; Volume 5, pp. 4929–4932.
22. Smith, S.C.; Seiler, P. Estimation with lossy measurements: jump estimators for jump systems. *IEEE Trans. Autom. Control* **2003**, *48*, 2163–2171. [CrossRef]
23. Matveev, A.S.; Savkin, A.V. The problem of state estimation via asynchronous communication channels with irregular transmission times. *IEEE Trans. Autom. Control* **2003**, *48*, 670–676. [CrossRef]
24. Lipsa, G.M.; Martins, N.C. Remote state estimation with communication costs for first-order LTI systems. *IEEE Trans. Autom. Control* **2011**, *56*, 2013–2025. [CrossRef]
25. Chakravorty, J.; Mahajan, A. Fundamental Limits of Remote Estimation of Autoregressive Markov Processes Under Communication Constraints. *IEEE Trans. Autom. Control* **2017**, *62*, 1109–1123. [CrossRef]
26. Nayyar, A.; Basar, T.; Teneketzis, D.; Veeravalli, V.V. Optimal Strategies for Communication and Remote Estimation With an Energy Harvesting Sensor. *IEEE Trans. Autom. Control* **2013**, *58*, 2246–2259. [CrossRef]
27. Sun, Y.; Polyanskiy, Y.; Uysal-Biyikoglu, E. Remote estimation of the Wiener process over a channel with random delay. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 321–325.
28. Linder, T.; Lugosi, G. A zero-delay sequential scheme for lossy coding of individual sequences. *IEEE Trans. Inf. Theory* **2001**, *47*, 2533–2538. [CrossRef]
29. Weissman, T.; Merhav, N. On Limited-Delay Lossy Coding and Filtering of Individual Sequences. *IEEE Trans. Inf. Theory* **2002**, *48*, 721–732. [CrossRef]
30. Weisman, T.; Merhav, N. Universal prediction of individual binary sequences in the presence of noise. *IEEE Trans. Inf. Theory* **2001**, *47*, 2151–2173. [CrossRef]
31. Matloub, S.; Weissman, T. Universal Zero-Delay Joint Source-Channel Coding. *IEEE Trans. Inf. Theory* **2006**, *52*, 5240–5249. [CrossRef]

32. A. K. Gorbunov, M.S.P. Nonanticipatory and Prognostic Epsilon Entropies and Message Generation Rates. *Probl. Inform. Transm.* **1973**, *9*, 184–191.
33. A. K. Gorbunov, M.S.P. Prognostic Epsilon Entropy of a Gaussian Message and a Gaussian Source. *Probl. Inform. Transm.* **1974**, *10*, 93–109.
34. Stavrou, P.; Kourtellaris, C.K.; Charalambous, C.D. Information Nonanticipative Rate Distortion Function and Its Applications. *CoRR* **2014**. [CrossRef]
35. Stavrou, P.A.; Østergaard, J.; Charalambous, C.D. Zero-Delay Rate Distortion via Filtering for Vector-Valued Gaussian Sources. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 841–856. [CrossRef]
36. Stavrou, P.A.; Charalambous, T.; Charalambous, C.D.; Loyka, S. Optimal Estimation via Nonanticipative Rate Distortion Function and Applications to Time-Varying Gauss–Markov Processes. *SIAM J. Control Optim.* **2018**, *56*, 3731–3765. [CrossRef]
37. Viswanathan, H.; Berger, T. Sequential coding of correlated sources. *IEEE Trans. Inf. Theory* **2000**, *46*, 236–246. [CrossRef]
38. Ma, N.; Ishwar, P. On Delayed Sequential Coding of Correlated Sources. *IEEE Trans. Inf. Theory* **2011**, *57*, 3763–3782. [CrossRef]
39. Khina, A.; Kostina, V.; Khisti, A.; Hassibi, B. Tracking and Control of Gauss–Markov Processes over Packet-Drop Channels with Acknowledgments. *IEEE Trans. Control. Netw. Syst.* **2019**, *6*, 549–560. [CrossRef]
40. Kipnis, A.; Reeves, G. Gaussian Approximation of Quantization Error for Estimation from Compressed Data. In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 2029–2033.
41. Delchamps, D.F. Extracting state information from a quantized output record. *Syst. Control. Lett.* **1989**, *13*, 365–372. [CrossRef]
42. Borkar, V.S.; Mitter, S.K., LQG Control with Communication Constraints. In *Communications, Computation, Control, and Signal Processing: A Tribute to Thomas Kailath*; Springer: Boston, MA, USA, 1997; pp. 365–373.
43. Wong, W.S.; Brockett, R.W. Systems with finite communication bandwidth constraints. II. Stabilization with limited information feedback. *IEEE Trans. Autom. Control* **1999**, *44*, 1049–1053. [CrossRef]
44. Nair, G.N.; Evans, R.J. Communication-limited stabilization of linear systems. In Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No. 00CH37187), Sydney, Australia, 12–15 December 2000; Volume 1, pp. 1005–1010.
45. Liberzon, D. On stabilization of linear systems with limited information. *IEEE Trans. Autom. Control* **2003**, *48*, 304–307. [CrossRef]
46. You, K.; Xie, L. Minimum data rate for mean square stabilization of discrete LTI systems over lossy channels. *IEEE Trans. Autom. Control* **2010**, *55*, 2373–2378. [CrossRef]
47. Yuksel, S. Stochastic stabilization of noisy linear systems with fixed-rate limited feedback. *IEEE Trans. Autom. Control* **2010**, *55*, 2847–2853. [CrossRef]
48. Yuksel, S.; Meyn, S.P. Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels. *IEEE Trans. Autom. Control* **2012**, *58*, 47–59. [CrossRef]
49. Yuksel, S. Characterization of information channels for asymptotic mean stationarity and stochastic stability of nonstationary/unstable linear systems. *IEEE Trans. Inf. Theory* **2012**, *58*, 6332–6354. [CrossRef]
50. Yuksel, S. Stationary and ergodic properties of stochastic nonlinear systems controlled over communication channels. *SIAM J. Control Optim.* **2016**, *54*, 2844–2871. [CrossRef]
51. Arnstein, D. Quantization error in predictive coders. *IEEE Trans. Commun.* **1975**, *23*, 423–429. [CrossRef]
52. Farvardin, N.; Modestino, J. Rate-distortion performance of DPCM schemes for autoregressive sources. *IEEE Trans. Inf. Theory* **1985**, *31*, 402–418. [CrossRef]
53. Gersho, A.; Gray, R.M. *Vector Quantization and Signal Compression*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 159.
54. Mayekar, P.; Tyagi, H. RATQ: A Universal Fixed-Length Quantizer for Stochastic Optimization. *arXiv* **2019**, arXiv:1908.08200.
55. Khina, A.; Khisti, A.; Kostina, V.; Hassibi, B. Sequential coding of Gauss-Markov sources with packet erasures and feedback. In Proceedings of the 2017 IEEE Information Theory Workshop (ITW), Kaohsiung, Taiwan, 6–10 November 2017; pp. 529–530.
56. Cover, T.; Thomas, J. *Elements of Information Theory*; Wiley: Hoboken, NJ, USA, 2006.