UNIVERSITY OF CALIFORNIA SAN DIEGO

LQG Control Performance under Coding Strategies in Network Control Systems

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Engineering Sciences (Mechanical Engineering)

by

Behrooz Amini

Committee in charge:

Professor Robert R. Bitmead , Chair
Professor Jorge Cortes
Professor Mauricio De Oliveira
Professor Bruce K. Driver
Professor Massimo Franceschetti

2020

The Dissertation of Behrooz Amini is approved and is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Chapter 3 is from the paper,“Predictive coding and control”, IEEE Transaction On Control of Network Systems, , Chun-Chia Huang, Behrooz Amini and Robert R. Bitmead 2018. Chapter 4 is from the paper,“LQG Control Performance with Low Bitrate Periodic Coding,” submitted to IEEE Transaction on Control of Network Systems, Behrooz Amini, Robert R. Bitmead 2020.

VITA

1996	Bachelor of Science, Mechanical Engineering, Isfahan University of Technology, Iran,
2004	Master of Science, Pure Mathematics, Amirkabir University of Technology, Iran,
2015	Master of Science, Mechanical Engineering, Southern Illinois University Edwardsville, U.S.A,
2015	Master of Science, Computational and Applied Mathematics, Southern Illinois University Edwardsville, U.S.A,
2020	Doctor of Philosophy, Engineering Sciences (Mechanical Engineering), University of California San Diego, U.S.A,

PUBLICATIONS

Chun-Chia Huang, **Behrooz Amini**, Robert R Bitmead. "Predictive coding and control", IEEE Transactions on Control of Network Systems, Volume 6, Issue 2, (2018) Page 906-918.

Behrooz Amini, Robert R Bitmead. "LQG Control Performance with Low Bitrate Periodic Coding", IEEE Transactions on Control of Network Systems, arXiv preprint arXiv:2004.03648, submitted 2020

Heinz Schättler, Urszula Ledzewicz, **Behrooz Amini**. "Dynamical properties of a minimally parameterized mathematical model for metronomic chemotherapy, Journal of mathematical biology, Volume 72, Issue 5, (2016) Page 1255-1280.

Urszula Ledzewicz, **Behrooz Amini**, Heinz Schättler. "Dynamics and control of a mathematical model for metronomic chemotherapy, Journal of Mathematical Biosciences and Engineering, Volume 12, Issue 6, (2015) Page 1257.

ABSTRACT OF THE DISSERTATION

LQG Control Performance under Coding Strategies in Network Control Systems

by

Behrooz Amini

Doctor of Philosophy in Engineering Sciences (Mechanical Engineering)

University of California San Diego, 2020

Professor Robert R. Bitmead , Chair

This thesis deals with a single feedback fixed-rate channel using some coding strategies. We assess and compare the LQ performance of the different coding methods. The idea of predictive coding is applied at the transmitter side to improve the efficiency of the channel usage by transmission of the quantized innovations signal. We observe a plant stability requirement is necessary to construct the joint density of both the plant and the predictor states at the receiver side. The Bayesian filter is used to compute the optimal feedback control. We compare the closed-loop control performance for three cases. In each of these competing cases, a lower complexity receiver architecture is possible but at the expense of closed-loop control performance.

In addition to predictive coding, we examine specific low-bitrate strategies and evaluate through their impact on LQ control performance. We consider coding the quantized output signal

deploying period-two codes of differing delay versus accuracy tradeoff. We treat the quantizer as the functional composition of an infinitely-long linear staircase function and a saturation. This permits the analysis being subdivided into estimator computations and an escape time evaluation, which connects the control back into the choice of quantizer saturation bound. By limiting the subject to specific strategies, we are able to identify principles underlying coding for control.

Chapter 1

Introduction

Over recent years, networked control problems and communication have gained much interest. In particular, the state estimation problem over a network has been widely studied. The problem of state estimation and stabilization of a linear time invariant system has been investigated with a number of differing communication constraints. Traditionally the areas of control and communication systems are studied separately as both have almost distinct underlying assumptions. For instance, in the area of control, one generally assumes perfect communication within the closed loop and receives data without time delay. On the other hand, in communication systems, data packets that carry the information can be delayed or dropped because of network traffic conditions. Further, and as studied in detail here, digital communications introduces quantization and channel noise. These different assumptions and natures create a barrier for researchers from the two fields to collaborate with each other. However, as new applications and technologies emerge, control and communication systems are shaping new horizons for more entwined and closer research to study.

As we see in some applications, observation and control signals are dispatched through a communication channel with a limited capacity or some bit-rate constraints. For instance, advances in large scale networks including sensors and actuators make an interesting area to explore. In sensor networks, the measurement data from various sensors are sent to the controller through a data network where data packets might be dropped or delayed if the network has severe traffic.

In this thesis, the architecture of control and communication systems under study is simply depicted in the Figure 1.1. We consider a general single digital channel for the study in which the physical system's arrangement is geographically distributed, say through the physical separation of the sensors from the plant.

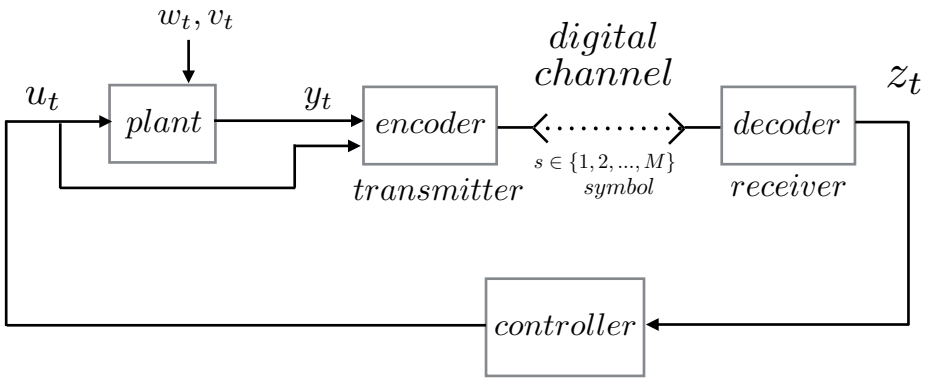

Figure 1.1. The network control system studied in this thesis.

We ideally would like to extend the current arrangement to have a secure ideal channel between the controller and plant. There is an underlying control objective of the control closed loop. This is distinct from many reliable communications objective of the communication subsystem, which might involve retransmission and devotion of bits to error detection and correct. We start with a closed-loop control performance objective function and apply optimal control concepts.

The arrangement of the control and communication systems consist of the following components.

- *Plant* is a linear time-invariant system subject to exogenous additive stochastic disturbances.

- The *encoder/transmitter* is a causal mapping from the measured output signal y_t. At the

transmitter, it determines the usage of the symbols sent through the channel and makes the efficient or robust use of the channel capacity. The process of coding is focused on efficient channel or bitrate, largely through removing redundancy in the signals.

- The *decoder/receiver* is required to be a causal mapping from the received quantized signal. It yields some reconstruction of y_t or the sequence of y_ts.

- The *codec*, encoder-decoder pair, is to facilitate the reconstruction of a close approximation of the transmitter-side signal by using the bits most effectively.

- The *Quantizer* (signal to symbol) is the simplest and delay free coder, which maps the current signal into one of the $M = 2^b$ symbols, where b is the number of bits per symbol. It is a mapping from the set of real value numbers as inputs to the finite set of outputs called output levels. A linear, midrise, eight-levels/three-bit quantizer with saturation is depicted in Figure 1.2. Linear here refers to the equal interval sizes assigned to each symbol.

  For m-vector signals or indeed for m-tuples of signals, this figure can be replaced by a partition of the $\mathbb{R}^m$ yielding one symbol per fundamental set. This latter construction is called a vector quantizer.

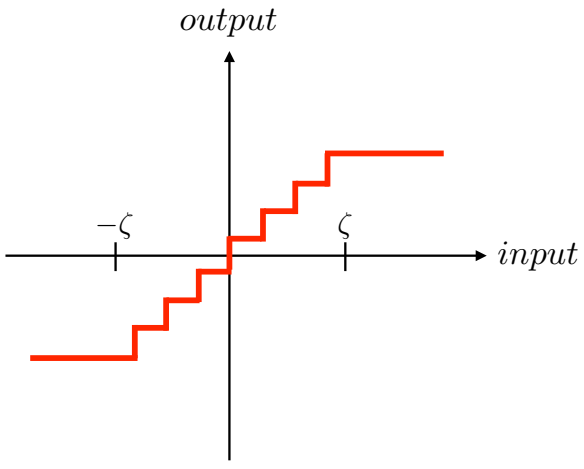

Figure 1.2. Graph of 3-bit/8-step staircase of a linear midrise quantizer-dequantizer pair with saturation at $\pm\zeta$.

Generally speaking quantization is a process to convert samples of a real value source into a digital representation with lower resolution. During this conversion, if we have too few bits or quantization levels, some information is lost. Loss of information maybe a reasonable concern, hence the quantization can be done such that the loss of information being imperceptible by proper designing and applying enough quantization levels. The primary goal is to make sure that the quantized signal is obtained at the proper resolution without harmful impact from the information is lost. The multi-element quantizers that we consider quantize a vector source input signal one element at each time. A quantizer, which resides at the transmitter, converts a real input signal into a symbol from a fixed alphabet. Then this symbol is transmitted through the channel, detected, and converted back into a real-values signal of the same dimension as the input vector. This latter reconstruction occurs at the receiver through a dequantizer.

- The *dequantizer* (symbol to signal) is the decoder associated with quantizer coder is called a dequantizer. Typically for a scalar quantizer the dequantized value on a receipt of symbol would be the midpoint of the interval mapping to a symbol at the quantizer. For vector quantization, one normally dequantizes by taking the centroid of the symbol's pre-image. When a quantizer and a dequantizer are combined, we have a mapping from real input signal at the transmitters, usual a signal taking a continuum of values, to a real output signal, which takes one of 2^b distinct values, where b is the bitrate. We shall adopt to convention to refer to the quantizer-dequantizer pair as the quantizer.

  In this thesis, we emphasize subtractive dithered quantization. Dithered quantization is a technique where we add a random signal to the input signal prior to quantization, and this makes the *quantization error* possess desirable statistical features as defined in Equation 1.1 where y is an input signal to the quantizer and z is an output signal from the quantizer.

$$\varepsilon = \mathbf{Q}(y) - y = z - y. \tag{1.1}$$

If the random dither signal is chosen properly, then the quantization error will be independent from the original input signal and, under certain conditions, will also be zero mean uniformly distributed white noise. In subtractive dither, a dither signal d_t is added prior to the quantizer input at the transmitter and subtracted from the dequantizer, output at the receiver as shown in Equation 1.2.

$$z_t = \mathbf{Q}(y_t + d_t) - d_t, \tag{1.2}$$

The goal is to have the quantization error independent from the plant output signal with favorable statistical properties as we explain in Chapter 2 in detail.

Regarding the quantization as a nonlinear memoryless measurement, Curry 's Theorem [1] plays a key role in our analysis. This theorem deals with the optimal feedback control of a linear system with memoryless nonlinear measurement and quadratic criterion function: the optimal controller is the LQ-optimal gain times the conditional state estimate.

Throughout the thesis quantizers will play a central role and in Chapter 2, we shall study quantization in considerable detail with an emphasis on linear mid-rise quantizers. We introduce and explore the idea of subtractive dithered quantization and establish properties of the approximation error between the quantizer input and output signals.

In Chapter 4, we also investigate three specific periodic coding strategies as period-two bit assignment and transmission strategies. The goal is to assess the efficacy of these coding strategies in terms of their advantages for quadratic-cost optimal output feedback control. For instance in Strategy I, we use a b-bit-per-channel subtractive dithered quantizer and send the b bits of the quantizer output through channel and at the receiver, use a Kalman filter to compute the filtered state estimate, $\hat{x}_{t|t}$, and control signal. In Strategy II, at even times we use a $2b$-bit-per-channel subtractive dithered quantizer and transmit the b most significant bits of the quantizer output through channel and at the receiver use the Kalman filter to calculate the filtered state estimate and control signal. At the odd times, we discard

the measurement signal and transmit the remaining b remaining bits of the quantized signal of even times to make the control signal by use of Kalman filter.

A major thrust of Chapter 4 will be to view the quantizer as the functional composition of an infinite staircase and a saturation. This perspective will allow us to tame the analysis by considering behavior prior to the first signal excursion beyond the saturation. Then separately we study the statistics of the first time to saturation, which we label the *escape time.*

- The communication *digital channel* (symbol-in symbol-out) is a physical system capable of transmitting a symbol from a finite alphabet and having that symbol received at the other end. In this thesis, we consider a fixed bitrate, error-free communications channel between the transmitter/plant side and the receiver/controller side. The sampling rate also will be fixed.

- The *controller* is based on optimal control and causally produces a control signal, u_t, from the received data. Throughout the thesis, we shall appeal to Curry's theorem for quadratic optimal control mentioned earlier.

In this thesis, similar to network control systems, we investigate methods such as predictive coding and other related approaches to achieve more efficient use of available bitrate, and therefore improves control performance. We focus on feedback control over a single fixed bit rate channel where in the transmitter side and apply predictor to make innovations signal similar to whitening filter prior to quantizer.

In Chapter 2, we present the quantization, quantization error and quantization bound and it is shown the details of input and output to far a midrise linear quantizer. We compare the quantization errors for infinite (unbounded) quantizer without saturation bound and the saturated case. For subtractive dithered quantizers, necessary and sufficient conditions are explored to have the probability density function of the quantization error being uniform. This condition has great practical results and consequence in the subsequent theorems of the thesis. The characteristic

function of the probability density has a crucial role in this chapter. We show that the subtractive dithered quantization error is a white noise under certain statistical features or conditions on the dither.

In Chapter 3 of this thesis we consider a very specific coder: linear predictive coding. Here, the transmitted symbol is a representation of the the prediction error or innovations signal at the plant output. This is produced with the aid of a Kalman filter running at the transmitter side. Such a symbol stream will be white in the Gaussian case and uncorrelated in general for linear plants. This leads to highly efficient usage of the channel bitrate. We evaluate a number distinct decoder options at the receiver: an innovations representation of the Kalman filter, and a Bayesian filter. Their distinction is a central contribution of the thesis. We provide a thorough evaluation of predictive coding when used as part of a network control system. In particular, we provide new theory on the role of predictive coding in stabilization and in performance. We show why we to assume that the open-loop plant system is stable in order to apply predictive coding, in particular for the linear Gaussian case.

In Chapter 3, we employ Bayesian filter to reconstruct the signal density at the receiver and we process the signal to compute an optimal control. Bayesian filter uses the sequence of measurements of control and innovations signal in the receiver side to compute the conditional density of the transmitter-side state. Bayesian filter offers the possibility of the computation which is not limit to Gaussian density and it is applicable to the general system having arbitrary probability density function. Hence, we do not have to restrict our computation to LQG performance index and therefore we can extend the cost function to Non-Lq optimal control due to the prowess of employing Bayesian filter at the receiver.

Then we apply feedback to assess the closed-loop performance objective by applying predictive coding. In addition, we consider some various simple coding strategies and compare the LQG costs. The predictive coding is applied to remove the redundancy of measurement signal and consequently fewer bits are required to communicate via the channel. The main purpose is to show optimal control based on predictive coding is improving the efficiency of the

channel usage.

We consider three comparative optimal control methods with different control signals for the closed-loop system in chapter four. The LQG cost functions were compared to apply control signal based on filtered and predicted state estimate from quantized innovations and filtered state estimate from quantized outputs. We will see the best performance pertains to the filtered state estimate from quantized innovations. As known the innovations remove the redundancy from the signal but it still keeps the information of the original measurement.

The role of full density reconstruction is established as central for nonlinear control rather than moment based methods. A major contribution is the proof, at least in the linear case, that feedback control based on transmission of the innovations sequence can not stabilize an unstable system. The unstable mode is not detectable. A nonlinear, non-quadratic optimal control problem is examined in detail. Advantages in closed-loop performance are demonstrated for plant state estimation based on quantized innovations versus quantized outputs or estimator state estimation.

In Chapter 4, regarding coding, our approach is inductive rather than deductive. This is a contrast with earlier works which, by and large, have been inconclusive. We treat a limited coding strategies in a way that we can make conclusions about the output signal. As we show as long as the controlled output signal is more predictable or correlated, we benefit from coding strategies. The range and correlation of the closed-loop output is crucial for the suitable choice of strategy. We have shown that if the controlled output signal is less predictable, Strategy I is more helpful and coding has less benefits, for instance those similar to minimum variance control. So the nature of signal plays a key role in picking the right strategy. In addition we see for the low bitrate the coding has significant impact, but its effects is diminishing by increasing the bit. The control objective function has a role to play in the efficacy of coding. This occurs because of its inherent effect on the predictability of the controlled output signal.

The main contribution of this chapter which differentiate from other research is treating the issue of signal escape time from the quantizer. We compute the residence time through

two methods and then compare the performance of coding strategies. The escape time analysis permits the consideration of stabilization problems and performance together. The focus on realization based behavioral descriptors admits new viewpoints compared with asymptotic moments. The escape time analysis paves the way to stabilize the system and computing performance simultaneously. Contribution is inductive more than deductive. This is a contrast with earlier works which, by and large, have been inconclusive.

1.1 Contributions

Dithered quantization

1. The seminal works of Gray and Stockham [2] and Widrow and Kollar [3] on dithered quantization are synthesized into a coherent whole with a focus on the role played by the saturation of the quantizer. These earlier authors establish the requirements of the subtractive dither signals and subsequent statistical properties of the quantization error. We profit from both aspects in our later studies.

2. But focusing on the consequences of saturation, we are able later in the thesis to concentrate on the behavior prior to the first saturation, which time we call the *escape time*. Further, we evaluate the probabilistic nature of the escape time as providing a finite time during which control performance can be evaluated using standard methods.

3. The central contribution of this chapter is to provide a thorough and consistent framework for the subsequent work. Technically, the new results extending the earlier works are otherwise minor.

Predictive coding and control

1. We provide a thorough evaluation of predictive coding when used as part of a network control system. In particular, we provide new theory on the role of predictive coding in stabilization and in performance.

2. The system and its predictor inhabit the transmitter side of the network. Since the predictive coder itself has a system model, the dynamics of the transmitter side are of dimension $2n$, where n is the system dimension. This leads to a rather non-obvious separation between the receiver-side reconstruction of the plant state versus of the predictor state. A Bayesian filter is developed to perform this reconstruction and a core contribution of the thesis is in demonstrating the control performance improvement from state estimation.

3. The role of the full density reconstruction is established as central for nonlinear control rather than moment based methods.

4. A major contribution is the proof, at least in the linear case, that feedback control based on transmission of the innovations sequence cannot stabilize an unstable system. The unstable mode is not detectable.

5. A nonlinear, non-quadratic optimal control problem is examined in detail. Advantages in closed-loop performance are demonstrated for plant state estimation based on quantized innovations versus quantized outputs or estimator state estimation.

LQG control performance with low bitrate periodic coding

1. Contribution is inductive more than deductive. This is a contrast with earlier works which, by and large, have been inconclusive.

2. The control objective function has a role to play in the efficacy of coding. This occurs because of its inherent effect on the predictability of the controlled output signal.

3. The prowess and novelties of this study help tackle the challenges arise from the quantizer 's saturation. We introduce the escape time first, evaluate the performance over that time. Then we decompose the quantizer into two stages - infinite levels quantizer and saturation - and it paves the way to consider the linear controlled covariances and escape time assessment simultaneously.

Chapter 2

Dithered quantization

2.1 Introduction

Waveforms are *continuous-time* and *continuous-amplitude* in nature, so *analog-to-digital* conversion is required to create a discrete representation of the waveform. Quantization is the key to analog-to-digital conversion and is inherently a non-linear mapping which may take any value from a large set as input (often continuous) to output values in a smaller set but often with a finite range. Before a signal can be processed by computer, its value must be sampled and quantized.

In this chapter, we investigate the conditions under which the error between a signal and its quantization version is uniformly distributed and we focus mostly on this quantization error. Since the quantization error is a deterministic function of input signal to the quantizer, we study the conditions which make the quantization error signal independent from the input signal.

First Widrow proved that if a random input signal has certain band-limited properties of characteristic function of probability density of the input signal, then the quantization error will be uniformly distributed. This requirement is actually a sufficient condition, but Sripad and Snyder [4] showed that uniform distribution can be achieved under a weaker condition which is actually necessary and sufficient. And also under this condition, they revealed [4] that the quantization error in two dimensional quantization channel are independent. By additional conditions, the error and input are uncorrelated but still they did not establish any conditions which imply independence of these signals.

In order to obtain the property that quantization error becomes independent of the input signal, we study dithered quantizers. A *dither* signal is added to the input signal before entering to the quantizer. This causes, under certain condition, the quantization error to be uniformly distributed, white and independent of the input signal.

The non subtractive dither was studied and developed by Stockham, Brinton, Lipshitz, Vanderkooy and Wannamaker, but subtractive dither is a clever idea suggested by Roberts (1962) to overcome the correlation properties of the quantizer. Then Schuchman in his paper [5] derived the conditions that a dither signal must meet such that the quantizer error is independent of the input signal for a finite level quantizer.

In 1993, Gray and Stockham published their paper [2] which was a thorough development of the conditions for random signals in general then they extended the results for input signals to the quantizers by normalizing the input and dither signals according to the quantization step. They have a clear and accurate idea by introducing two conspicuous types of quantization errors for non subtractive and subtractive errors, ε and ε' respectively. And they found the necessary and sufficient conditions for which the signal input is being independent of quantization errors pertinent to the input signal for non subtractive and subtractive quantizers.

We eventually come across QTSD theorem [3] which is introduced by Widrow. The theorem reveals the sufficient condition for which the quantization error is independent of the input signal but it is still incomplete. We will introduce an enriched version of the QTSD theorem considering other pioneers' papers in the field and elaborate the necessary and sufficient conditions that the quantization error being independent of input signal.

2.2 Quantizer Definition

A scalar quantizer is a map $\mathbf{Q} : \mathbb{R} \to \mathscr{L}$, where $\mathscr{L}$ is a finite ordered or countably infinite ordered set, but our focus in this thesis is on finite levels with saturation bound which we explain

shortly. Let us consider the set

$$\mathscr{L} = \{l_1, l_2, l_3, ..., l_M\} \subset \mathbb{R},$$

of real numbers such that

$$l_1 < l_2 < ... < l_M.$$

The set $\mathscr{L}$ is referred to as the alphabet. An M-level quantizer $\mathbf{Q}$ is defined by a set of $2M-1$ strictly ordered real numbers: $M-1$ breakpoints $\{y_1, \ldots, y_{M-1}\}$ defining intervals

$$\begin{aligned} Q_1 &= (-\infty, y_1] \\ Q_2 &= (y_1, y_2] \\ &\vdots \\ Q_M &= (y_{M-1}, +\infty) \end{aligned}$$

and M levels $\{l_1, \ldots, l_M\}$ jointly satisfying the interlacing property

$$y_0 = -\infty < l_1 < y_1 < l_2 < y_2 < \cdots < y_{M-1} < l_M < y_M = +\infty,$$

such that $\mathbf{Q}(y) = l_i$ if $y \in Q_i$. If $M = 2^b$ for integer b, then the indices, i, for the l_is can be stored in b-bits. We define the bit rate to be b bits/sample, where

$$b = \log_2 M \quad \text{bits/sample.}$$

This set of indices i provide an efficient method to store quantized values and is referred to as a codebook, since the quantizer output can be produce from the table of l_i values. The indices are also referred to as symbols. A *uniform* or *linear quantizer* is one where saturation bounds are defined $-\zeta < l_1$ and $l_M < \zeta$ such that the intervals $[-\zeta, y_1], (l_1, l_2], \ldots, (l_M, \zeta]$ are of equal

length and l_is are the midpoints of the interval containing it.

In this thesis, we shall limit discussion to scalar uniform quantizers with integer bitrate; these are defined by the two quantities ζ and b, the saturation bound and bitrate. It is possible to treat vector quantizers, where the input is a vector signal and output is a scalar index decoded into a vector of the original dimension. We shall not consider general vector quantizers, which are discussed in [6], but treat quantization of vectores as a vector of scalar quantizers, one in each dimension or channel.

A quantizer can be considered as the combined operations of an encoder and a decoder, jointly called the coder. The encoder is a mapping $\mathscr{Q} : \mathbb{R} \to \mathscr{S} = \{1,2,3,...,M\}$, where $\mathscr{S}$ is the set of symbols, and the decoder is the mapping $\mathscr{Q}^{-1} : \mathscr{S} \to \mathscr{L}$, such that if $(\mathscr{Q}^{-1}\mathscr{Q})(y) = l_i = z$ then $\mathscr{Q}(y) = i$ and $\mathscr{Q}^{-1}(i) = l_i$. As we see the structure of *coder* or quantizer in the following Figure,

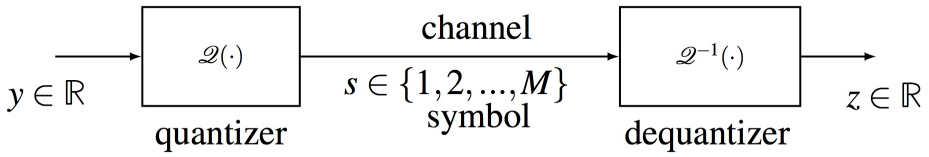

Figure 2.1. Schematic of coder/quantizer.

The actual mapping $z = \mathbf{Q}(y)$ is a staircase function with unity slope shown in Figure 2.2 for uniform quantizer. There are several types of quantizers but we are interested in the midrize uniform quantizer. For instance in Figure 2.2, a three-bit or eight-level quantizer is shown with

quantizer bound $[-\zeta, \zeta]$ where $\zeta = 4$,

$$\Delta = \frac{2\zeta}{2^b} = \frac{\zeta}{2^{b-1}}.$$

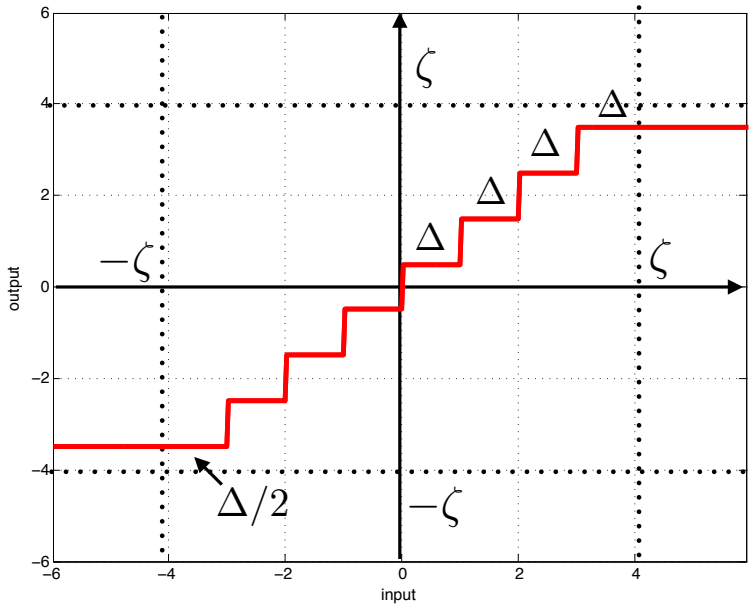

Figure 2.2. 8-level/3-bit midrise quantizer, step size $\Delta = 1, \zeta = 4$.

In many applications, the number of levels M is chosen to be very large, so the quantized output is a very close approximation to the original input. If the quantizer has infinite but countable output levels, M is not finite number, then we call quantizer an *infinite* quantizer which does not have saturation bound. We shortly explain the difference between these quantizers.

2.2.1 Quantization Error

Quantization error or round-off error is the difference between an input value and its output (quantized value).

$$\varepsilon = \mathbf{Q}(y) - y = z - y. \tag{2.1}$$

The quantization error ε is a nonlinear memoryless function of the input signal y. In case of uniform quantizer with a bounded input, for instance

$$y \in (-\zeta, \zeta),$$

the quantization *step size* is

$$\Delta = \frac{2\zeta}{2^b} = \frac{\zeta}{2^{b-1}}. \tag{2.2}$$

where quantization errors have the values in the range

$$\Delta/2 \leq \varepsilon < \Delta/2.$$

If we assume the probability density function of the quantization errors has distributed uniformly in the above range,

$$p_\varepsilon(y) = \begin{cases} \frac{1}{\Delta} & |\varepsilon| \leq \frac{\Delta}{2} \\ 0 & \text{otherwise} \end{cases}$$

then the variance of quantization errors $\varepsilon = \mathbf{Q}(y) - y = l_i - y$, can be calculated by changing variable as follows

$$\begin{aligned} \sigma_\varepsilon^2 &= \int_{-\infty}^{\infty} (\mathbf{Q}(y) - y)^2 p_Y(y) dy, \\ &= \int_{-\infty}^{\infty} \varepsilon^2 \frac{1}{\Delta} d\varepsilon, \\ &= \frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \varepsilon^2 d\varepsilon, \end{aligned}$$

so the variance of the quantization error that is uniformly distributed for an interval of width Δ from the Equation 2.2 is

$$\sigma_{\varepsilon}^{2} = \frac{\Delta^{2}}{12} = \frac{\zeta^{2}}{3 \times 2^{2b}}. \tag{2.3}$$

Equation (2.3) is helpful to see the the standard deviation of the error increases by step size and also it decreases exponentially by increasing the number of quantizer's level.

2.3 Linear Quantizers

Depending on the type of quantizer, finite or infinite level quantizer, the quantizer errors have different structures. For instance, we compare infinite level quantizer and a 3-bit quantizer for the input interval $[-2, 2]$ in the following.

In Figure 2.3 an infinite quantizer is shown, as we see the quantizer error signal is a periodic function of input signal *y* because it does not saturate. But in the 3-bit quantizer, Figure 2.4, the saturation happens and we see the quantization error increases once the input signal jumps out of the quantizer saturation bound $[-1, 1]$. Both types of error have exact quantization error within the saturation bound $[-1, 1]$ such that the difference between finite level quantizer and infinite quantizer level is zero within the bound $[-1, 1]$, but the quantization errors vary outside of the saturation bound.

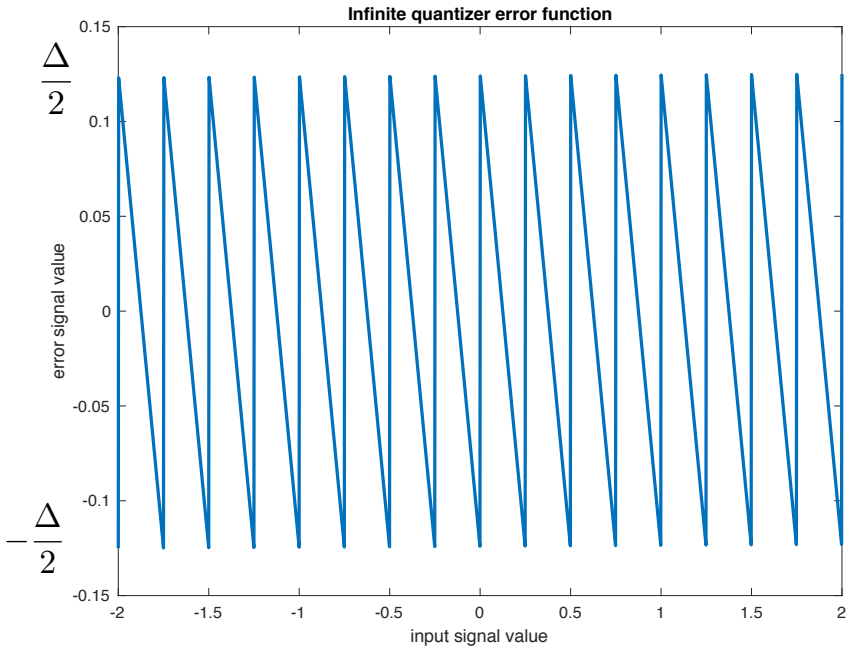

Figure 2.3. Periodic quantizer error for infinite quantizer without saturation bound.

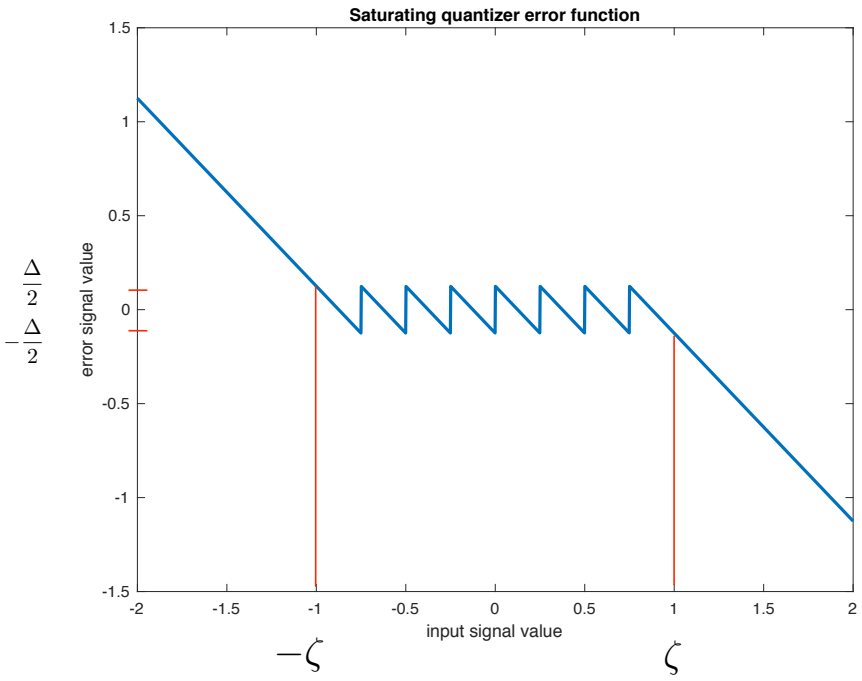

Figure 2.4. Quantizer error for 3-bit quantizer with saturation input bound $[-1, 1]$.

2.3.1 Linear Infinite Level Quantizer

Let us now consider an infinite midrize quantizer as depicted in Figure 2.3, we see the quantizer error is a periodic function of the input signal to the quantizer, so we can formulate the probability density function of the quantization error ε as a function of the probability density function of input signal y.

$$p_\varepsilon(\varepsilon) = \begin{cases} \sum_{n=-\infty}^{\infty} p_Y(n\Delta+\varepsilon), & -\Delta/2 \le \varepsilon < \Delta/2 \\ 0, & \text{otherwise.} \end{cases}$$

In this section, a necessary and sufficient condition is given such that the probability density function of the quantization error is uniform. This was proved by Sripad and Snyder [4].

Theorem 1. *[4] The probability density of the quantization error of an infinite uniform is*

$$p_\varepsilon(\varepsilon) = \begin{cases} \frac{1}{\Delta} + \frac{1}{\Delta} \sum_{n\neq 0} \Phi_Y(\frac{2\pi n}{\Delta}) \exp(\frac{-j2\pi n\varepsilon}{\Delta}), & -\Delta/2 \le \varepsilon < \Delta/2 \\ 0, & \textit{otherwise} \end{cases} \tag{2.4}$$

where Φ_Y is the characteristic function of the random input variable Y to the quantizer.

Proof: The probability density $p_\varepsilon(\varepsilon)$ can be denoted in terms of the density $p_Y(y)$ according to

$$p_\varepsilon(\varepsilon) = \begin{cases} \sum_{m=-\infty}^{\infty} p_Y(m\Delta+\varepsilon), & -\Delta/2 \le \varepsilon < \Delta/2 \\ 0, & \text{otherwise.} \end{cases}$$

By defining

$$g(\varepsilon) = \sum_{m=-\infty}^{\infty} p_Y(m\Delta+\varepsilon), \quad -\Delta/2 \le \varepsilon < \Delta/2,$$

and we see $g(\varepsilon)$ is periodic with period of Δ, on $[-\Delta/2, \Delta/2)$, but we can easily extend the

periodicity from finite interval $[-\Delta/2, \Delta/2)$ to $(-\infty, \infty)$.

Since $g(\varepsilon)$ is periodic with a period Δ, so it can be represented by a Fourier series

$$g(\varepsilon) = \sum_{n=-\infty}^{\infty} \tilde{g}(n) \exp\left(\frac{j2\pi n\varepsilon}{\Delta}\right) \tag{2.5}$$

where $\tilde{g}(n)$ are the Fourier coefficients and are computed as

$$\begin{aligned}\tilde{g}(n) &= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \sum_{m=-\infty}^{\infty} p_Y(m\Delta+\varepsilon)\exp\left(\frac{-j2\pi n\varepsilon}{\Delta}\right) d\varepsilon \\ &= (1/\Delta)\phi_Y(-2\pi n/\Delta), \quad n = 0, \pm 1, \pm 2, ...\end{aligned}$$

by substituting the Fourier coefficients in (2.5) and use of the property $\Phi_Y(0) = 1$, then obtains

$$g(\varepsilon) = \frac{1}{\Delta} + \frac{1}{\Delta}\sum_{n\neq 0} \Phi_Y\left(\frac{-2\pi n}{\Delta}\right)\exp\left(\frac{j2\pi n\varepsilon}{\Delta}\right). \square$$

Corollary 2. *[4] The density function of the quantization error for an infinite quantizer is uniform, i.e.,*

$$p_\varepsilon(\varepsilon) = \begin{cases} 1/\Delta, & -\Delta/2 \leq \varepsilon < \Delta/2 \\ 0, & \text{otherwise,} \end{cases} \tag{2.6}$$

if and only if the characteristic function of the input random variable satisfies

$$\Phi_Y(2\pi n/\Delta) = 0 \quad \text{for all} \quad n \neq 0. \tag{2.7}$$

In the case of having two random variables as the input to a two-dimension of infinite uniform quantizer, we have the following theorem that shows the quantizer errors are independent from each other.

Theorem 3. *[4] The joint density of the quantization error for an infinite quantizer is uniform; i.e.,*

$$p_{\varepsilon_1,\varepsilon_2}(\varepsilon_1,\varepsilon_2) = \begin{cases} 1/\Delta^2, & -\Delta/2 \le \varepsilon_1 < \Delta/2, -\Delta/2 \le \varepsilon_2 < \Delta/2 \\ 0, & \textit{otherwise,} \end{cases} \tag{2.8}$$

if and only if the joint characteristic function of the input random variables satisfies

$$\Phi_{X_1,X_2}\left(\frac{2\pi l}{\Delta}, \frac{2\pi k}{\Delta}\right) = 0 \quad \textit{for all} \quad l \neq 0 \quad \textit{and} \quad k \neq 0 \tag{2.9}$$

Remark: If (2.9) is satisfied, then the two quantization errors ε_1 and ε_2 in each channel are independent and distributed uniformly, through (2.6) and (2.8), we obtain

$$p_{\varepsilon_1,\varepsilon_2}(\varepsilon_1,\varepsilon_2) = p_{\varepsilon_1}(\varepsilon_1) p_{\varepsilon_2}(\varepsilon_2).$$

2.3.2 Linear Finite-Level Quantizer

Now let us apply general idea of Theorem 1 to a finite level quantizer when *no saturation* occurs, then we have the following theorem for M-level quantizer.

Theorem 4. *Consider an M-level uniform quantizer with input signal y where no saturation occurs. Then the probability density of the quantization error is*

$$p_\varepsilon(\varepsilon) = \begin{cases} \frac{1}{\Delta} + \frac{1}{\Delta} \sum\limits_{n \neq 0} \Phi_Y(\frac{2\pi n}{\Delta}) \exp(\frac{-j2\pi n\varepsilon}{\Delta}), & -\Delta/2 \le \varepsilon < \Delta/2 \\ 0, & \textit{otherwise} \end{cases} \tag{2.10}$$

where Φ_Y is the characteristic function of the random input variable Y to the quantizer.

We can exactly apply Corollary 2 for a finite level quantizer provided that it does not saturate.

2.4 Examples

Quantization error behaves as function of input signal. In the following we consider signal $y = \sin t$ as input to 3-bit and 4-bit quantizers, and we compare the quantization error between both. Note that the structure and magnitude of the quantization error ε for input signal $y = \sin t$ are similar in both Figures 2.5 and 2.6, but the quantization error in Figure 2.6 is with an amplitude of one quarter the value of ε in Figure 2.5. In addition we clearly realize the quantization errors depend on the input signal or we see deterministic relation between these.

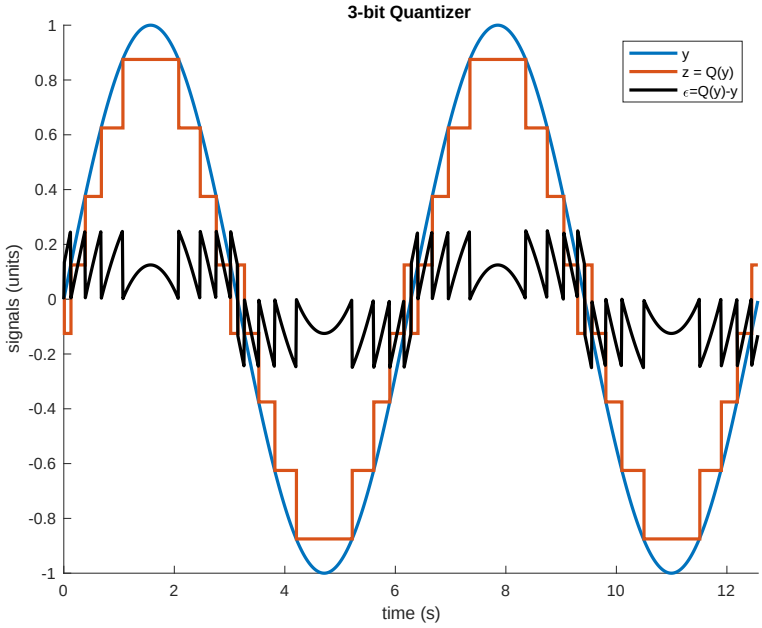

Figure 2.5. Quantizing $\sin x$ with 3-bit .

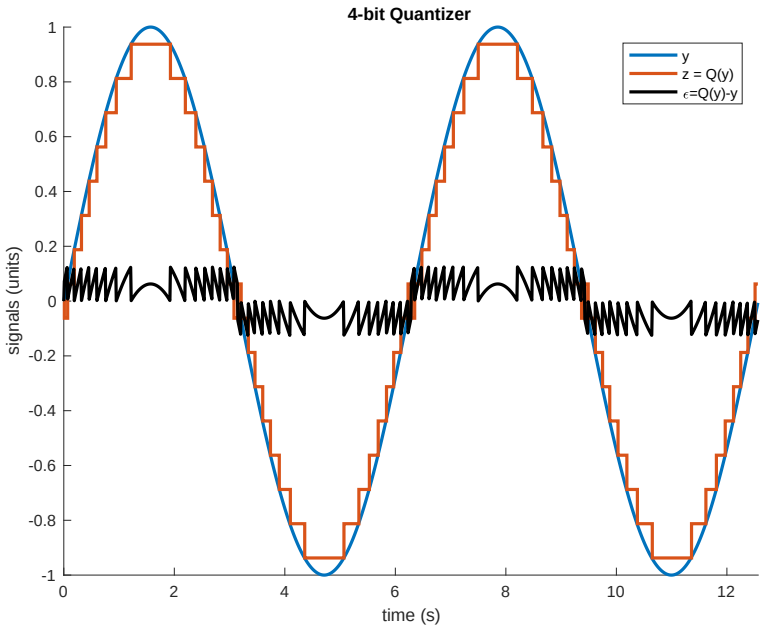

Figure 2.6. Quantizing $\sin x$ with 4-bit .

The signal $y = \sin t$ which is quantized with the 3-bit previously, now is studied in the Figures 2.7 and 2.8, we see the autocorrelation of quantization error and cross-correlation of quantization error. Both Figures show clear correlation between the error signal and between it and the signal $y = \sin t$ being quantized.

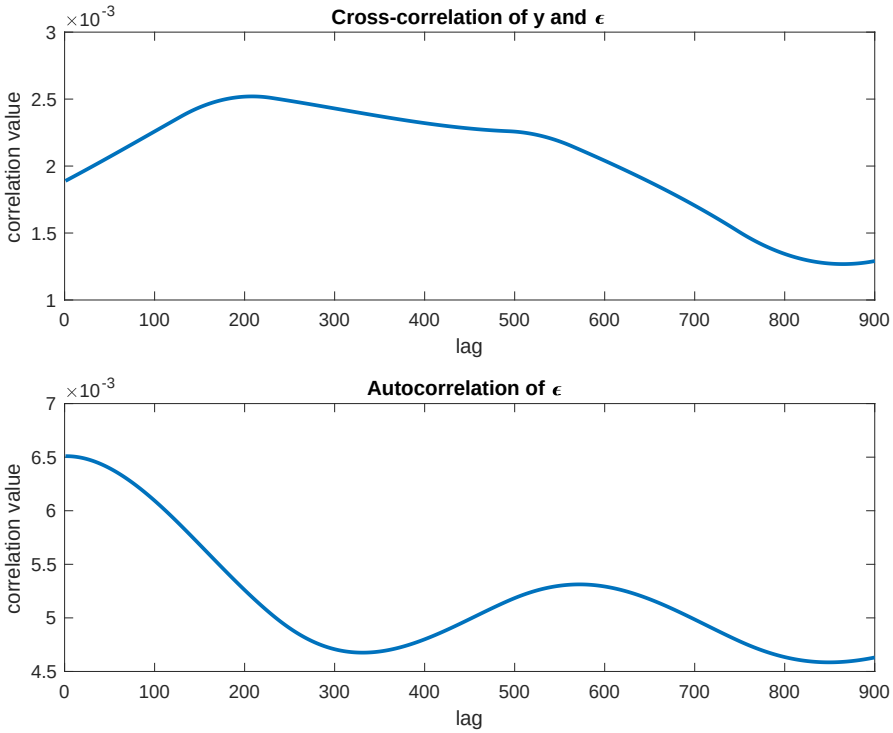

Figure 2.7. Crosscorrelation (top) and autocorrelation (bottom) where $\sin t$ is the input to 3-bit quantizer.

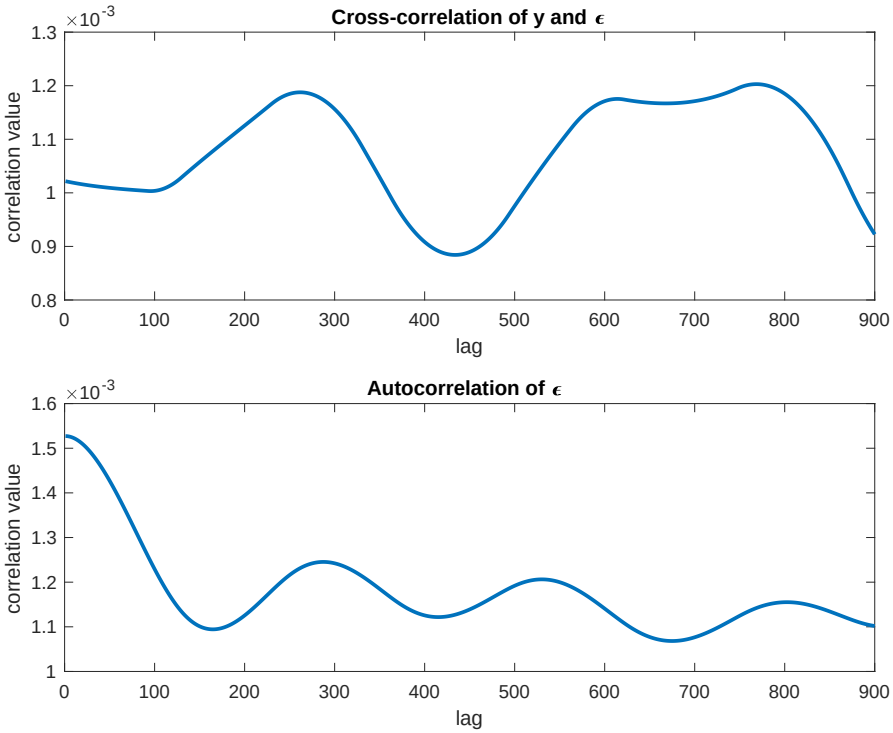

Figure 2.8. Crosscorrelation (top) and autocorrelation (bottom) where $\sin t$ is the input to 4-bit quantizer.

In the previous examples which do not saturate, we observe these

- the quantization error ε is deterministic function of signal y
- as the number of quantizer 's level M increases for the same ζ, the quantization error ε decreases
- the quantization error is not a white noise signal
- $\varepsilon \in [-\Delta/2, \Delta/2]$
- the quantization error ε and input signal y are correlated.

For the rest of this chapter, the apex of our goal is to look for desired properties of the quantization error which has the following features

1. The signal $\{\varepsilon_t\}$ be white and zero mean.
2. The signals $\{y_t\}$ and $\varepsilon_{t+\tau}$ are uncorrelated and orthogonal, $E(y_t\varepsilon_{t+\tau}) = E(y_t)E(\varepsilon_{t+\tau}) = 0$ for all t and $\tau \neq 0$.
3. The signal $\{\varepsilon_t\}$ has a uniform probability density distribution.[6]

We address later the conditions under which the above properties being achieved. The output of the quantizer can be computed as the input to the quantizer plus quantization error

$$\mathbf{Q}(y_t) = y_t + \varepsilon_t.$$

In the next section, *subtractive dither* is introduced. Dither is a random signal that it is added to the input signal prior to quantization to randomize the quantization error. By applying subtractive dither, under several conditions, the quantization error becomes independent of the signal being quantized and quantization error has uniform density.

2.5 Subtractive Dithered Quantizer

As it can be seen in Figure 2.9, a dither is a random signal that is added to the quantizer input, and subtracted it from the quantizer output. This requires the receiver to have the same dither as transmitter.

Theorem 5. *[2] Assume a dither signal d is independent of the input signal y and the quantizer does not overload,* $|y+d| \leq \zeta$. *Then the condition*

$$\Phi_d(\frac{2\pi n}{\Delta}) = 0, \quad n \neq 0, \tag{2.11}$$

is necessary and sufficient for achieving the following properties

- *signal input y is independent of the quantizer error*

$$\varepsilon = \mathbf{Q}(y+d) - (y+d).$$

- *The quantizer error* ε *is distributed uniformly on* $[-\Delta/2, \Delta/2]$.

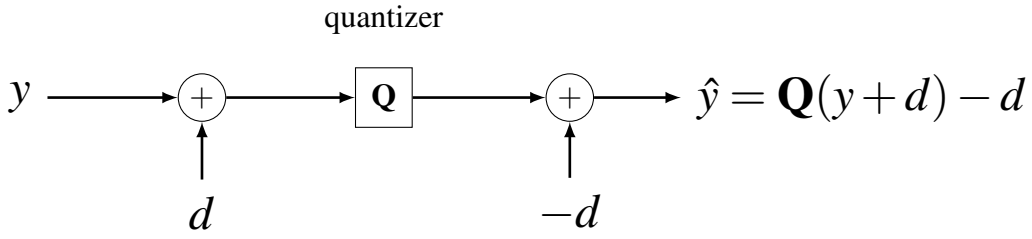

Figure 2.9. [Subtractive dithered quantizer.]

In the following we see how apply the theorem for two different dithers with uniform and triangular densities.

Characteristic function with uniform density:

Consider dither with uniform density on $[-a,a]$.

$$p_{d,a}(x) = \mathscr{U}[-a,a] = \begin{cases} \frac{1}{2a}, & x \in [-a,a] \\ 0, & \text{else.} \end{cases} \tag{2.12}$$

Then let us find the characteristic function or Fourier transform of this density,

$$\begin{aligned} \Phi_{d,a}(\omega) &= \mathscr{F}(p_{d,a}(x)), \\ &= \int_{-\infty}^{\infty} p_{d,a}(x) e^{jx\omega} dx, \\ &= \int_{-a}^{a} \frac{e^{jxw}}{2a} dx, \\ &= \frac{1}{-j2a\omega}[e^{-ja\omega} - e^{ja\omega}], \\ &= -\frac{1}{-j2a\omega} \times -2j \times \sin a\omega, \\ &= \frac{1}{a\omega}\sin(a\omega) = \text{sinc}\, a\omega, \end{aligned}$$

so we obtain,

$$\Phi_{d,a}\left(l\frac{2\pi}{\Delta}\right) = \text{sinc}\, l2\pi\frac{a}{\Delta}.$$

If $a = \frac{\Delta}{2}$ then $\Phi_d(l\frac{2\pi}{\Delta}) = \text{sinc}\, l\pi = \begin{cases} 1, & l = 0, \\ 0, & l = \pm 1, \pm 2, ... \end{cases}$

If $a = \Delta$ then $\Phi_d(l\frac{2\pi}{\Delta}) = \text{sinc}\, 2l\pi = \begin{cases} 1, & l = 0, \\ 0, & l = \pm 1, \pm 2, ... \end{cases}$

This suggests that we should pick the smallest support for dither which is $\mathscr{U}[\frac{-\Delta}{2}, \frac{\Delta}{2}]$. In the following examples later, we apply the uniform dither which is depicted in the Figure 2.10.

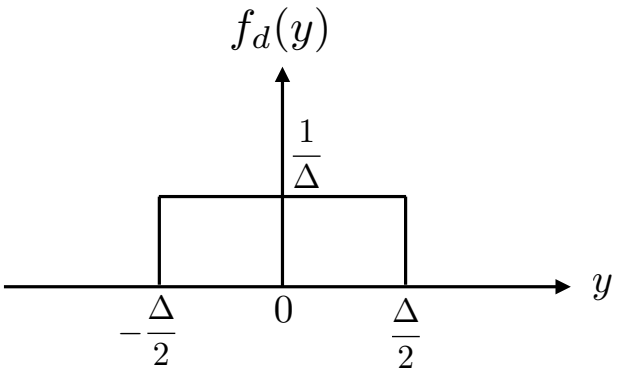

Figure 2.10. Dither signal with uniform density

Characteristic function with triangular density:

If we sum up two identical uniform densities properly normalized and centered then we have triangular density. The triangular density is the convolution of two uniform densities and its Fourier transform is the product of the Fourier transforms and so satisfies the zero property as said in above theorem.

Consider $v = u_1 + u_2$ where $u_1, u_2 \sim \mathscr{U}[-a,a]$, then the pdfs convolve

$$\begin{aligned}
p_v(x) &= p_{u_1}(x) * p_{u_2}(x), \\
&= \frac{1}{4a^2}\int_{-\infty}^{\infty} p_{u_1}(\tau)p_{u_2}(x-\tau)d\tau, \\
&= \frac{1}{4a^2}\begin{cases} 0, & x < -2a \\ \int_{-a}^{x+a} d\tau, & -2a < x < 0, \\ \int_{x-a}^{a} d\tau, & 0 < x < 2a, \\ 0, & 2a < x, \end{cases} \\
&= \frac{1}{4a^2}\begin{cases} 0, & x < -2a \\ x+2a, & -2a < x < 0, \\ 2a-x, & 0 < x < 2a, \\ 0, & 2a < x, \end{cases} \\
&= \mathrm{tr}[-2a, 2a].
\end{aligned}$$

the fourier transform relation also shows that

$$\Phi_v(\omega) = \Phi_u(\omega)^2.$$

So if $\mathscr{U}[\frac{-\Delta}{2}, \frac{\Delta}{2}]$ dither satisfies QTSD, then so does $\mathrm{tr}[-\Delta, \Delta]$ dither but $\mathrm{tr}[\frac{-\Delta}{2}, \frac{\Delta}{2}]$ does not.

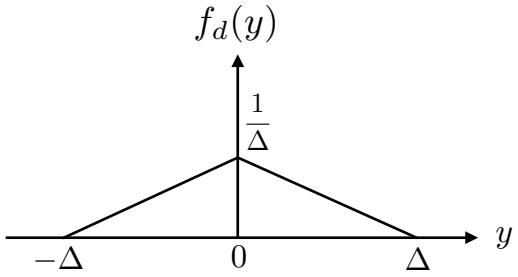

Figure 2.11. Dither signal with triangular density

2.6 Refined Version of QTSD Theorem and Quantizing Time Series

A *subtractive b-bit dithered quantizer,* $\mathbf{Q}(\cdot)$, is a memoryless function which takes input signal y_t and dither signal, d_t, and produces an output signal

$$z_t = \mathbf{Q}(y_t + d_t) - d_t, \tag{2.13}$$

(2.14)

which is shown in Figure 2.12

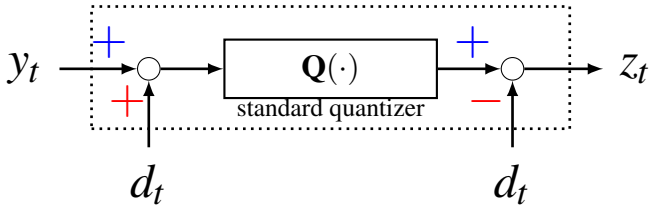

Figure 2.12. [Subtractive dithered quantization: $\mathbf{Q}(\cdot)$]

and the quantization error is defined

$$\varepsilon_t = z_t - y_t = \mathbf{Q}(y_t + d_t) - y_t - d_t, \tag{2.15}$$

where $Q(\cdot)$ is a standard quantizer. Such quantizers are examined in detail in, for example, [3].

The QTSD theorem (Quantizing Theorem for Subtractive Dither) shows sufficient conditions to make the quantization error zero-mean white uniform distribution independent of the signal being quantized.

Theorem 6. *Consider a linear, midrise, symmetric, b-bits-per-channel, subtractive dithered quantizer,* $\mathbf{Q}(\cdot)$*, with saturation bounds* $\pm\zeta$ *in each channel. Assume:*

(A) dither d_t *is a stationary white noise process independent from* y_t *with a probability density possessing characteristic function* $\Phi_d(\cdot)$*,*

(B) $y_t + d_t \in [-\zeta, \zeta]$*, i.e. no saturation of the dithered quantizer occurs.*

Then, the quantization error

$$\varepsilon_t \triangleq \mathbf{Q}(y_t + d_t) - (y_t + d_t), \tag{2.16}$$

is white, independent from y_t*, uniformly distributed on* $\left[-\frac{\zeta}{2^b}, \frac{\zeta}{2^b}\right] = [-\frac{\Delta}{2}, \frac{\Delta}{2}]$*,*

if and only if

$$\Phi_d\left(\frac{2\pi l}{\Delta}\right) = \Phi_d\left(l\frac{\pi 2^b}{\zeta}\right) = 0 \quad \textit{for} \quad l = \pm 1, \pm 2 \ldots. \tag{2.17}$$

Denote the quantizer step size as

$$\Delta = \frac{\zeta}{2^{b-1}},$$

and

$$\begin{aligned} &\varepsilon_t \sim \mathscr{U}\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right], \ \mathrm{E}(\varepsilon_t) = 0, \\ &\mathrm{cov}(\varepsilon_k) = \frac{\zeta^2}{3 \times 2^{2b}} \triangleq S_b. \end{aligned} \tag{2.18}$$

Proof: Let us first prove that we can show the joint characteristic function satisfies

$$\Phi_{\varepsilon_n,Y_k}(u,v) = \Phi_{\varepsilon_n}(u)\Phi_{Y_k}(v),$$

we apply the nested expectation

$$\Phi_{\varepsilon_n,Y_k}(u,v) = E\left[e^{juY_k}E\left[e^{jv\varepsilon_n}|Y_n,Y_k\right]\right], \tag{2.19}$$

since d_n is independent from Y_n and Y_k, the conditional expectation evaluated at $Y_n = y_n$ and $Y_k = y_k$ is given by

$$E\left[e^{ju\varepsilon_n}|Y_n = y_n, Y_k = y_k\right] = E\left[e^{ju\varepsilon_n}\right].$$

Now for a fixed y_n, the given condition (2.17) imply ε_n is uniformly distributed on $[-\Delta/2, \Delta/2]$. Then we have

$$E[e^{ju\varepsilon_n}|Y_n,Y_k] = \frac{\sin(u\Delta/2)}{u\Delta/2} = \Phi_U(u) \tag{2.20}$$

and by applying the equation (2.19) we obtain

$$\Phi_{\varepsilon_n,Y_k}(u,v) = E\left[e^{juY_k}\right]\Phi_U(ju) = \Phi_{Y_k}(u)\Phi_U(u). \tag{2.21}$$

Equation (2.20) also implies that

$$\Phi_{\varepsilon_n}(u) = E[E[e^{j\varepsilon_n}|Y_n,Y_k]] = \Phi_U(u). \tag{2.22}$$

From equations (2.21) and (2.22) both imply ε_n is uniformly distributed on $[-\Delta/2, \Delta/2]$ for all n and ε_n and Y_k are independent for all n and k.

Now let us show ε_n is uniformly distributed and ε_n and ε_l are independent for any $l \neq n$.

We have

$$\begin{aligned}\Phi_{\varepsilon_n,\varepsilon_l}(u,v) &= E[e^{ju\varepsilon_n+jv\varepsilon_l}] \\ &= E[E[e^{ju\varepsilon_n+jv\varepsilon_l}|Y_n,Y_l]].\end{aligned}$$

But we for given $Y_n = y_n$ and $X_l = x_l$ the random variables ε_n and ε_l are conditionally independent, since d_n and d_l are independent and are both uniformly distributed. Thus the conditional expectation is

$$E[e^{ju\varepsilon_n+jv\varepsilon_l}|Y_n,Y_l] = \Phi_U(u)\Phi_U(v) = \Phi_{\varepsilon_n}(u)\Phi_{\varepsilon_l}(v).$$

Theorem 6 is an embellishment of Theorem QTSD of [3] and theorem 5 present *necessary and sufficient* conditions under which the quantization error is an additive white noise independent from the signal being quantized. We also, note that the characteristic function condition is satisfied by dither which is uniform $\mathscr{U}[-\Delta/2,\Delta/2]$ or which is triangularly distributed tr$[-\Delta,\Delta]$ for example.

Let us apply subtractive dither to sine wave and reevaluate auto-covariance of quantization error in Figures 2.13 and 2.14 where these are 3-bit and 4-bit quantizers respectively. We see in the subtractive dither the quantization error is decreased and also it is white. By increasing the bits, we see the quantization error is decreased as well.

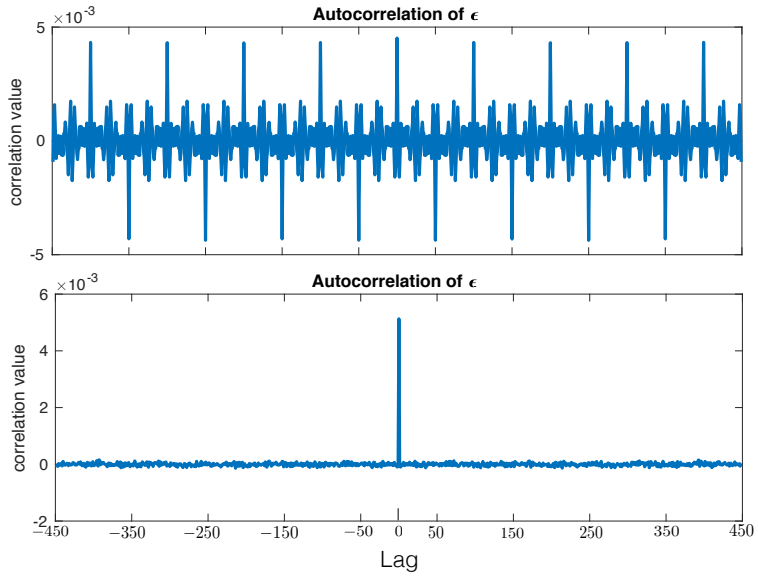

Figure 2.13. $\mathrm{E}(\varepsilon(t+\tau)\varepsilon(t))$ without SD (top) and with SD (bottom), 3-bit quantizer.

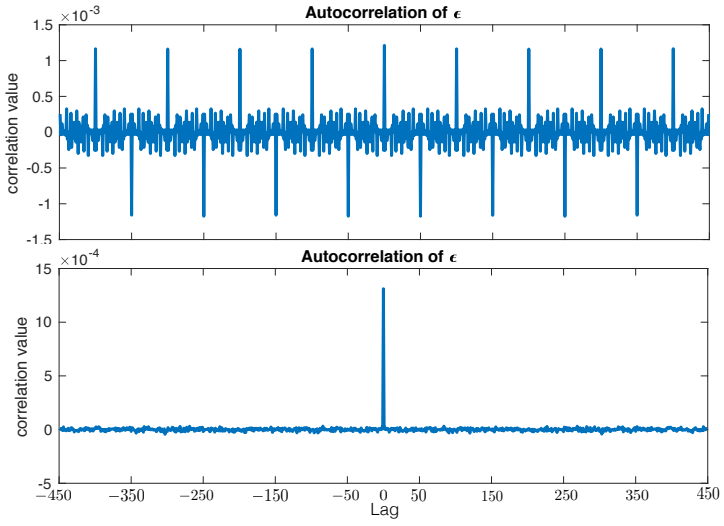

Figure 2.14. $\mathrm{E}(\varepsilon(t+\tau)\varepsilon(t))$ without SD (top) and with SD (bottom), 4-bit quantizer.

Now let us investigate the cross correlation between the input signal and the quantization error and compare two cases, with subtractive dither and without subtractive dither. As we see in Figure 2.15 the quantization error ε_1 is dependent on the input signal y.

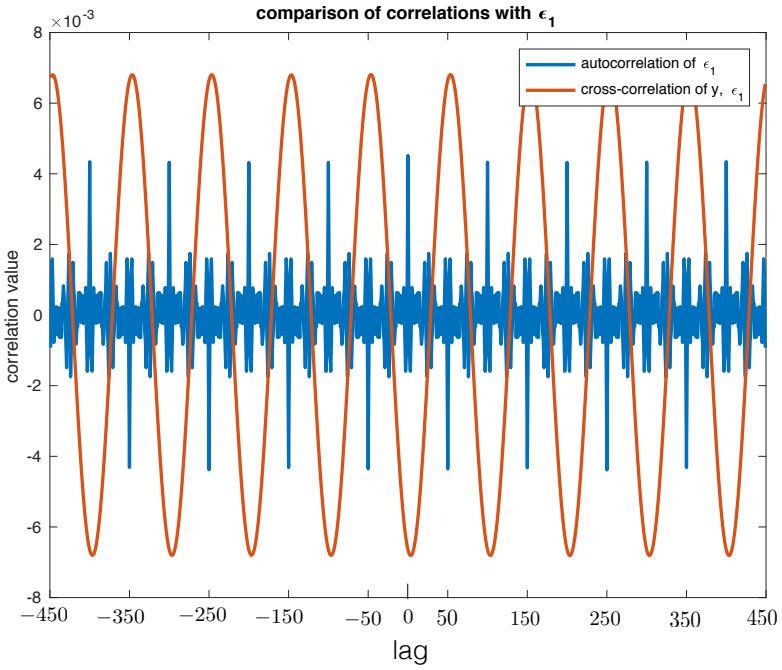

Figure 2.15. Autocorrelation of ε_1 and cross-correlation of y and ε_1.

The correlation between the quantization error and the input signal can be seen in the Figure 2.16. But the autocorrelation of quantization error is reduced significantly as it is shown in the Figures 2.15 and 2.16. In the Figure 2.17 we can compare two cross-correlations where both show the quantization errors is dependent on the input signal but the subtractive case has less cross-correlation.

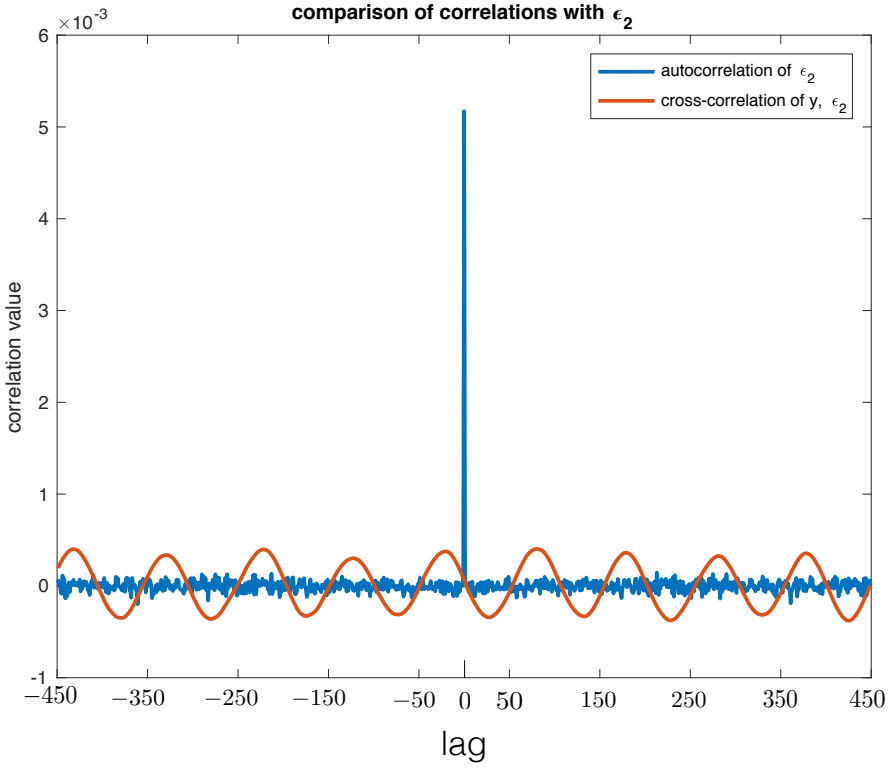

Figure 2.16. Autocorrelation of ε_2 and cross-correlation of y and ε_2.

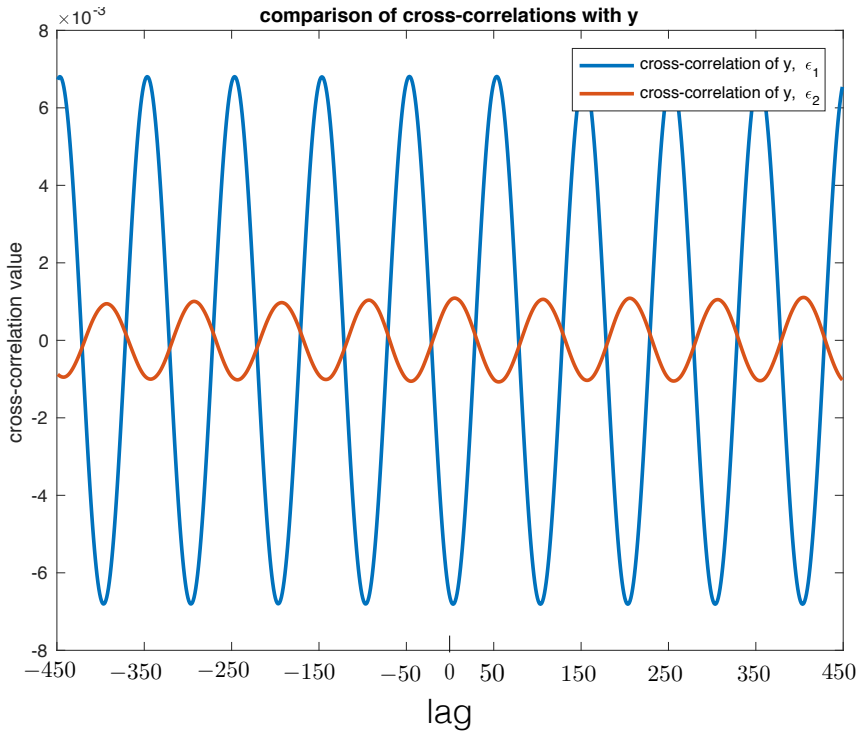

Figure 2.17. Cross-correlation of y and ε_1 and cross-correlation of y and ε_2.

We need to answer that why there is still correlation for subtractive dither case between the input signal and quantization error. In order to address this issue we should check whether the conditions in theorem 6 are satisfied or not. We know the signal input is $Y_t = \sin t$, so the density of the signal which is deterministic signal is $f_Y(y) = \delta(y_t - \sin t)$, hence

$$\Phi_y(Y) = \int_{-\infty}^{\infty} \delta(y(t) - \sin t) e^{juy} dy = e^{ju \sin t}$$

and it is clearly does not satisfy the condition 2.7. So we can not apply Theorem 6. We saw in the previous example a nonstationary signal $y(t)$, where its characteristic function is a function of time and also its being deterministic signal.

In the Figure 2.18, we can apply the result of QTSD Theorem for an example with strongly correlated random y(t), say a very low-pass filtered white noise process, and we observe that QTSD Theorem perfectly works.

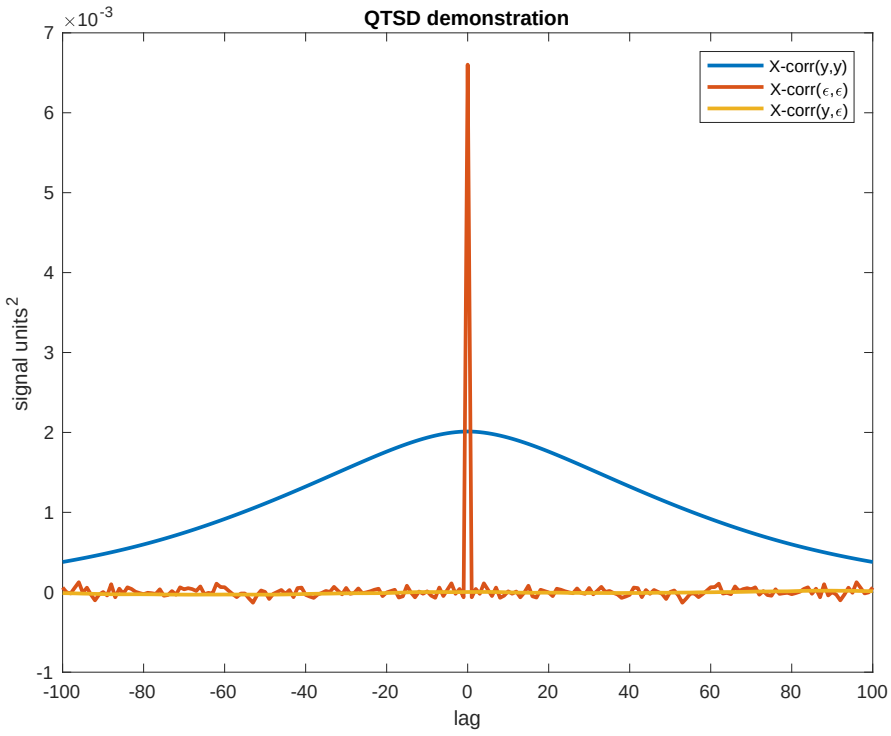

Figure 2.18. Autocorrelation of y, autocorrelation of ε and cross-correlation of y and ε.

Chapter 3

Predictive coding and control

3.1 Abstract

This chapter deals with feedback control over a single fixed-rate channel using predictive coding at the transmitter side. The central thrust is to demonstrate that optimal control based on predictive coding plus fixed memoryless quantization at the transmitter, designed to improve the efficiency of the channel usage and exemplified (or perhaps extremized) by the transmission of the quantized innovations signal, in general requires the construction of the joint density of both the plant and predictor states at the receiver side and inherits a plant stability requirement, which is examined. The Bayesian filter is developed. This recursive filter's state density is used to compute the optimal feedback control. This is in contrast to the less complicated propagation solely of the predictor state, which would suffice in the linear quadratic optimal control problem – a feature that is elucidated. A linear non-quadratic optimal control example is provided to illustrate the approach and its benefits over control based on the recovered predictor state density or control without predictive coding. In each of these competing cases, a lower complexity receiver architecture is possible but at the expense of closed-loop control performance.

3.2 Introduction

Predictive quantization [6] is used in communication systems to whiten the transmitted digital signal and remove redundancy, thereby improving coding performance. In delay-free

coding environments, a prediction of the source signal is computed and then subtracted from the signal to yield a prediction error, which is then quantized. Compared with the original signal, the prediction error is both closer to white, i.e. less correlated over time, and possesses a smaller variance, which aids in scaling the quantizer range for improved effectiveness. Such systems form the basis of familiar schemes such as ITU-T G.721/722/726 Adaptive Differential Pulse Coded Modulation (ADPCM) standards [7] and Delta Modulation [6]. The ADPCM schema is depicted in Figure 3.1. The quantizer $\mathcal{Q}$ is fixed and the adaptive predictor and gain serve to whiten and limit dynamic range fluctuations of the transmitted error signal, thereby improving distortion between transmitter and receiver. The decoder/receiver mimics the encoder/transmitter

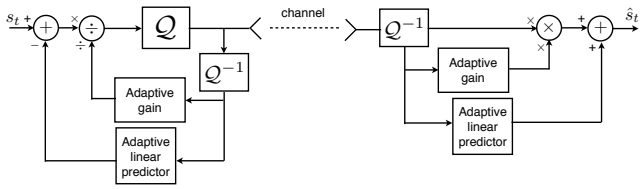

Figure 3.1. ITU-G.722 Adaptive Differential Pulse-Coded Modulation schema.

to undo its operations and recover an approximation, $\hat{s}_t$, of the signal s_t. ADPCM has been proven in service in telecommunications systems since at least 1984 and has spawned a number of variants as commercial lossy speech compressors. ADPCM has been interpreted as a disturbance rejection feedback control system in [8]. It provides a kindred example in the paper, but without its attendant gain adaptation, which could bring it closer to [9], nor its limitation to using the receiver-side signal at the transmitter. It manifests similar stability requirements.

For network control systems, these methods can be applied in the link between plant and controller to achieve more efficient use of the available link bit-rate and, thereby, improved control performance because the more effective coding leads to more accurate reconstruction of the transmitted signal at the receiver. It is this reconstruction and, in particular, plant state density estimation, which is the focus of this paper. The Bayesian filter is used to calculate the joint and marginal densities of the plant and predictor states conditioned on the received data. The general set-up is depicted in Figure 3.2 and will be made precise shortly.

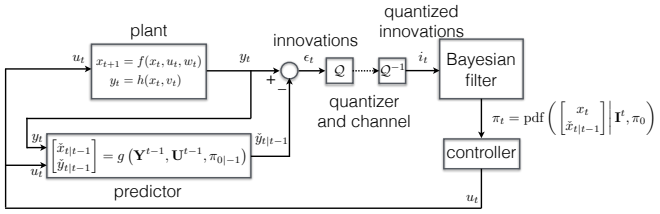

Figure 3.2. Predictive quantization based feedback control set-up.

While the structure and analysis with the Bayesian filter pertains for general nonlinear systems, the most edifying and best studied case concerns linear plants with Gaussian noise paired with the Kalman filter as the state and output predictor. In this case, the innovations sequence is Gaussian, zero mean and white. Quantized innovations state estimation has been studied in this case in [10, 11, 12, 13], both from the perspective of state estimation and of control, notably LQG control. We too shall specialize to the linear Gaussian case, since the analysis is both relatively direct and most informative, because the comparator unquantized controls are so well known. However, we shall take an immediate departure from the linear predictive coding approach of signal processing by seeking at the receiver to compute the precise conditional state density, $p(x_t|\mathbf{I}^t, \pi_0)$, rather than to capture the transmitter side prediction or output, $\check{x}_{t|t-1}$ or $\check{y}_{t|t-1}$. Indeed, part of the message is that the plant state density is in general the important aspect for control; a point which we illustrate with an example.

Pertinent prior literature

Quantization in control under communications constraints is a longstanding subject with emphases on both stabilization and performance [13, 14, 15, 9]. Studies include adaptive quantization of the control signal and of the plant output measurement sequence and include simultaneous coding and quantization. A subset of papers studies *dynamic quantization* [16, 17] in which the coding is restricted to a finite-dimensional linear system. Almost universally, the setting is linear systems control with a quadratic criterion function. From a signal processing and telecommunications perspective, predictive coding [6] is familiar and involves pre-whitening of the signal before quantization. In a Kalman filtering framework, such methods equate to

quantization of the innovations sequence, which inherently is white, Gaussian with minimum variance. Thus, the innovations is optimally coded and so we seek a static memoryless quantizer as in ADPCM. This is a restriction that we make on the coder.

Beginning with Fischer [18] the connection between quantization and LQ control performance was studied with the optimal quantizer being time-varying, although with an asymptotically time-invariant dequantizer, with level boundaries depending on the current state estimate. For Fischer, the quantizer operates on the computed optimal control signal. Fu [19] and Yüksel [20] extend these results to causal coding quantizers and Fu identifies some technical errors in [18]. The focus is on fixed-rate quantizer design and the existence or otherwise of a separation theorem in this case. The work [21] treats a related output feedback control problem with variable bit rate coding and random channel delay, deriving bounds for the limiting average codeword length given a specified bound on the controlled state covariance.

Papers [22, 23] consider the *cost-rate tradeoff* in linear quadratic control. This problem seeks the lowest average bitrate, $\mathbb{R}(b)$, channel required to achieve a specified LQ performance b. Stavrou et al. [24] treat a related Kalman filtering problem and seek the minimal data rate required to achieve a specific distortion or mean squared error between the plant state and the receiver-side Kalman filter. All three papers start from a vector autoregressive plant model with fully measured state at the transmitter. They restrict attention to zero-delay coding schemes which is appropriate for feedback control. Each of these papers arrives at a coding scheme based on Shannon entropy coding of a quantized innovations, but a different innovations from here.

Each of [22, 23, 24] and ourselves has a single bandlimited forward channel and a high-fidelity return channel, used in [24] for communication of the receiver state estimate and in [22, 23] and here to communicate the control signal. Kostina and Hassibi [22] and Tanaka et al. [23] require stabilizability of the plant system's $[A,B]$ pair and adjust the coding to accommodate the plant feedback stabilization as part of their calculation. This is evident in their inherent satisfaction of Tatikonda's and Mitter's [25] and Nair's and Evans' [9] lower bound on the bit rate based on the unstable eigenvalues of A. Here, because we quantize directly the output

innovations process constructed at the transmitter, we must limit the analysis to stable plants. This is a property proven in Corollary 10 and explained as it arises in ADPCM. Tanaka et al. [23] and Stavrou et al. [24] incorporate the communication of the receiver's state estimate of the plant state to compute their state innovations, which is then quantized in the coder. We note too that the encoding strategy in [22] also is based on quantizing then entropy coding the state innovations between the true plant state and the receiver's estimate and can accommodate partial state measurements with Gaussian noise. Initial values, channel noise and quantization error force the state estimates at the transmitter and receiver to differ. The effect of this is seen in the additional stability condition in Corollary 10. To improve performance and simplify analysis the quantizers can be subtractively dithered in each of these works.

Uniform subtractive dithered quantizers of finite support have been analyzed in [26, 27], where they demonstrated that such quantizers and predictive coding arise in achieving a Gaussian nonanticipative rate distortion function with a specified mean square filtering error overbound, provided quantizer overload is avoided. In this context, they treat full-state transmission over n-parallel AWGN channels with feedback of the state prediction from the decoder. They also propose an approach to mitigate the effects of overload. This formulation differs from ours in the full-state communication and in the feedback of the receiver prediction. The computation of the innovations process before encoding, however, is similar and demonstrated to be close to optimal for their constrained zero-delay coding problem. For us, we take the predictive coder with uniform subtractive dithered quantizers in each channel as the starting point. Papers [26, 27] provide a justification of this as a sensible starting point. The work of [28] connects some of these coding aspects to the feedback control problem.

3.2.1 Contributions

- The results commence from the noisy plant output measurement rather than the full state and extend to nonlinear systems with non-quadratic optimal control. The treatment includes quantized LQG and computed examples.

- They expose the role of the Bayesian filter and state conditional density, rather than moment, reconstruction.

- They are based on or limited to predictive coding at the transmitter using quantized output innovations by a memoryless, fixed-rate quantizer. This solution is necessarily zero-delay and fixed bit rate, in comparison with the other entropy coded approaches which are variable rate.

- Our problem focus is to optimize the plant performance given the communications structure based on predictive coding. This is in contrast to [22, 23, 24] where the control performance is specified and communications required to achieve this is then designed.

A number of papers are dedicated to reconstruction of the conditional density of either the plant state or the predictor state using methods allied with Kalman filtering [10], Bayesian filtering [11] and particle filtering [12]. These are closest in focus to the current work, although they are limited to the consideration of linear systems. Once the transmitter-side prediction and quantization scheme is decided, the problem that we consider is the reconstruction by Bayesian filter at the receiver of the filtered conditional density of the plant state, as opposed to the density of the predictor state. We do this in a fully nonlinear context and then specialize to the linear problem. We provide theory and demonstrate by example the control performance benefits of using: the plant state density, the filtered density versus the predicted density, and the quantized prediction error versus quantized output signals. The contribution is to provide a unifying nonlinear framework in which to treat the predicted and filtered state conditional density reconstruction and to explore its connection with other approaches from the linear context based on reconstruction of the predictor state conditional density or its mean value.

Notation

We denote probability density functions (pdfs) by $p(\cdot)$. Gaussian pdfs of mean μ and covariance P are denoted $\mathscr{N}(\mu, P)$. The initial pdf of the transmitter state will be denoted $\pi_{0|-1}$.

The data available at the receiver at time t is $\mathbf{I}^t = \{\pi_{0|-1}, i_0, \ldots, i_t\}$. By the same token, the data available at the transmitter is $\mathbf{E}^t = \{\pi_{0|-1}, \varepsilon_0, \ldots, \varepsilon_t\}$. We presume that, at time t, the input signal, u_t, computed at the receiver/controller is available also to the transmitter side.

3.3 Nonlinear Predictive Quantization – Transmitter side

We consider separately the general case of a nonlinear plant at the transmitter side and its specialization to a linear Gaussian system. The Bayesian filter construction applies to both but the linear formulation allows us to draw on well understood ideas from Kalman filtering. For comparison and brevity, we present them side by side. Although the quantized linear innovations problem has been more widely studied.

3.3.1 Nonlinear plant & predictor

The nonlinear stochastic plant system is described by

$$x_{t+1} = f_t(x_t, u_t, w_t), \quad x_0, \tag{3.1}$$

$$y_t = h_t(x_t, v_t). \tag{3.2}$$

Here, state $x_t \in \mathbb{R}^{n_x}$, input $u_t \in \mathbb{R}^{n_u}$, output $y_t \in \mathbb{R}^{n_y}$, process noise $w_t \in \mathbb{R}^{n_w}$, measurement noise $v_t \in \mathbb{R}^{n_v}$. Noise sequences $\{w_t\}$ and $\{v_t\}$ are mutually independent, zero-mean and white with known densities. The plant initial condition, x_0, has known density, $\pi_{0|-1}$, and is independent from w_t and v_t for all t.

The measured output and control signals at the transmitter, u_t and y_t, are the inputs to a finite-dimensional predictor

$$\xi_{t+1} = \bar{g}_t(\xi_t, u_t, y_t), \quad \xi_0, \tag{3.3}$$

$$\check{y}_t = j_t(\xi_t). \tag{3.4}$$

The prediction, in turn, is combined with y_t to produce a prediction error or *innovations* signal.

$$\varepsilon_t = y_t - \check{y}_t. \tag{3.5}$$

Using (3.4)-(3.5), then (3.3) becomes

$$\xi_{t+1} = g_t(\xi_t, u_t, \varepsilon_t), \quad \xi_0, \tag{3.6}$$

since y_t can be reconstructed from ε_t and $\check{y}_t$.

3.3.2 Linear Gaussian plant & predictor

The linear plant system is described by

$$x_{t+1} = Ax_t + Bu_t + w_t, \tag{3.7}$$

$$y_t = Cx_t + v_t, \tag{3.8}$$

where $\{w_t\}$ and $\{v_t\}$ are mutually independent, white noises of known densities and also zero-mean Gaussian with covariances Q and R respectively. The state estimator and predictor is the Kalman predictor with state $\check{x}_t$ and innovations

$$\varepsilon_t = \begin{bmatrix} C & -C \end{bmatrix} \begin{bmatrix} x_t \\ \check{x}_t \end{bmatrix} + v_t. \tag{3.9}$$

The predictor recursion is

$$\begin{aligned}
\check{x}_{t+1} &= A\check{x}_t + Bu_t + L_t\varepsilon_t, && (3.10)\\
&= (A - L_tC)\check{x}_t + Bu_t + L_ty_t,\\
&= (A - L_tC)\check{x}_t + L_tCx_t + Bu_t + L_tv_t, && (3.11)\\
\check{y}_t &= C\check{x}_t,\\
L_t &= A\Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + R)^{-1},
\end{aligned}$$

where

$$\begin{aligned}
\Sigma_{t|t-1} &= E\{[x_t - \check{x}_{t-1}][x_t - \check{x}_{t-1}]^T|\mathbf{E}^{t-1}\},\\
\Sigma_{t|t} &= E\{[x_t - \check{x}_t][x_t - \check{x}_t]^T|\mathbf{E}^{t-1}\}.
\end{aligned}$$

Here the covariance matrix, $\Sigma_{t|t-1}$, is given by the Riccati difference equation commencing from $\Sigma_{0|-1}$ [29]. The prediction error or *innovations* is given by (3.9).

3.3.3 Quantization

The 'quantizer' (or combined quantizer-dequantizer pair) in Figures 3.1 and 3.2, is a known single-valued function of the same dimension, n_y, as its input. Typically (and in our calculations below), the quantization function, $\mathscr{Q}$, maps real intervals to unique fixed digital signal values in each channel and the dequantization function, $\mathscr{Q}^{-1}$, maps these received values to unique points in their corresponding intervals.

$$i_t = \mathbf{Q}(\varepsilon_t) = \mathscr{Q}^{-1}(\mathscr{Q}[\varepsilon_t]). \tag{3.12}$$

Gersho and Gray [6] describe many quantizer designs for communication systems, covering both scalar and vector quantization including optimization for properties such as minimal distortion and the Lloyd-Max quantizer, which is adapted to signals with Gaussian

distributions.

Assumption 1. *The quantizer-dequantizer function,* $\mathbf{Q}(\cdot)$*, is known, finite range and memoryless.*

Our formulation makes no other specific assumptions about the quantizer. Although, our calculations later for linear systems are based on a uniform (linear) mid-rise quantizer and its mid-point 'inverse'. We arbitrarily absorb the quantizer into the transmitter side, since we only care about the quantizer-dequantizer pair in the signal domain without regard to the specifics of the channel representation. Although, it is straightforward to incorporate other features into the full formulation, such as channel noise or the ADPCM structure as above.

More generally, the quantizer is a codec, a coder-decoder pair. Since we focus on control, we limit our study to delay-free coding. We consider memoryless coding for simplicity in order not to have to incorporate the codec states and to exploit the whiteness of the innovations. The important feature of quantized innovations signal, $\{i_t\}$, compared with the transmitter-side innovations, $\{\varepsilon_t\}$, is that it has reduced information content, measured by entropy or other metrics, and this therefore diminishes its utility as a means to compute the optimal control. Understanding the cost of quantization borne by the control performance in this setting is the aim of this paper.

The following result, a specialization of the Data Processing Inequality [30], captures this relationship.

Lemma 7. *Denote the following* σ*-algebras:* $\mathscr{I}_t = \sigma(\mathbf{I}^t)$ *and* $\mathscr{E}_t = \sigma(\mathbf{E}^t)$*. Then, since* $i_t = \mathbf{Q}(\varepsilon_t)$ *for known* $\mathbf{Q}(\cdot)$*,*

$$\mathscr{I}_t \subseteq \mathscr{E}_t. \tag{3.13}$$

Subtractive dithered quantizer for the linear case

A subtractive dithered quantizer, $\mathscr{Q}_d(\cdot)$, consists of a fixed, finite-range quantizer function $\mathfrak{Q}(\cdot)$ with a predetermined dither signal, $\{d_t\}$, which is known to both transmitter and receiver.

It is defined

$$\mathscr{Q}_d(\varepsilon_t) \stackrel{\triangle}{=} \mathfrak{Q}(\varepsilon_t + d_t) - d_t. \tag{3.14}$$

A *linear* quantizer is one with equally spaced steps with the center points of the steps mapping the input value to the same output value. A *midrise* quantizer has discontinuity at the origin of the input. We have this result following [3].

Theorem 8. *Suppose quantizer* $\mathbf{Q}(\cdot)$ *is a subtractive, dithered, b-bit-per-channel, midrise, symmetric, linear quantizer with saturation values* $\pm\zeta$*. Suppose, further, that the subtractive dither signal is white, independent in each channel, and either uniformly distributed* $\mathscr{U}(-\zeta/2^b, \zeta/2^b)$ *or triangularly distributed,* $d_t \sim tr(-\zeta/2^{b-1}, \zeta/2^{b-1})$*, and known exactly to the transmitter and receiver. If the signal* $\varepsilon_t + d_t \in [-\zeta, \zeta]$*, then the quantization noise*

$$\psi_t \stackrel{\triangle}{=} i_t - \varepsilon_t = \mathscr{Q}_d(\varepsilon_t) - \varepsilon_t, \tag{3.15}$$

is white, independent from $\{\varepsilon_t\}$*, and uniformly distributed* $\psi_t \sim \mathscr{U}[-\zeta/2^b, \zeta/2^b]$*. That is,*

$$E(\psi_t) = 0, \;\; E(\psi_t^2) = \frac{\zeta^2}{3 \times 2^{2b}} \stackrel{\triangle}{=} \Psi. \tag{3.16}$$

3.3.4 Transmitter assumptions

Assumption 2. *1.* $x_t \in \mathbb{R}^{n_x}$, $u_t \in \mathbb{R}^{n_u}$, $y_t \in \mathbb{R}^{n_y}$, $\xi_t \in \mathbb{R}^{n_\xi}$.

2. *Control signal* u_t *is known to both transmitter and receiver at time* $t+1$.

3. $\{w_t\}$ *and* $\{v_t\}$ *are mutually independent, white noises of known densities and, in the linear case, also zero-mean Gaussian with known covariances.*

4. *The transmitter-side initial states,* x_0 *and* ξ_0 *(respectively* $\check{x}_0$*), have known joint density,*

$\pi_{0|-1}$, *independent from* $\{w_t, v_t\}$. *In the linear case,*

$$\pi_{0|-1} = \mathscr{N}\left(\begin{bmatrix}\hat{x}_{0|-1}\\ \hat{x}_{0|-1}\end{bmatrix}, \begin{bmatrix}\Sigma_{0|-1} & 0\\ 0 & 0\end{bmatrix}\right).$$

The receiver has knowledge of these densities.

5. *In the nonlinear case, if the conditional mean of the plant state is computed at the transmitter, it is a function of* ξ_t.

$$\check{x}_t = E(x_t|\mathbf{Y}^{t-1}) = E(x_t|\mathbf{E}^{t-1}) = \ell_t(\xi_t). \tag{3.17}$$

6. *The function* $g_t(\cdot,\cdot,\cdot)$ *in* (3.6) *causes the predictor state update to be uniformly incrementally input-to-state stable [31]. In the linear case,* A *in* (3.7) *and* (3.10) *has all eigenvalues strictly inside the unit circle. The origin of this stability condition, at least in the linear case, is examined in detail in Subsection 3.4.1.*

3.4 Quantized Innovations Bayesian Filtering – Receiver Side

The nonlinear signal model for the sequence, $\{i_t\}$, arriving at the receiver comprises:

- For the nonlinear case,
  - Using (4.2), (3.4) and (3.5), $\varepsilon_t = h_t(x_t, v_t) - j_t(\xi_t)$. Then (3.6) and (3.7) yield the

combined state recursion

$$
\begin{aligned}
z_{t+1} &\stackrel{\triangle}{=} \begin{bmatrix} x_{t+1} \\ \xi_{t+1} \end{bmatrix} = \begin{bmatrix} f_t(x_t, u_t, w_t) \\ g_t(\xi_t, u_t, \varepsilon_t) \end{bmatrix}, \\
&= \begin{bmatrix} f_t(x_t, u_t, w_t) \\ g_t(\xi_t, u_t, h_t(x_t, v_t) - j(\xi_t)) \end{bmatrix}, \\
&\stackrel{\triangle}{=} \mathfrak{f}_t(z_t, u_t, w_t, v_t). \qquad (3.18)
\end{aligned}
$$

– Output equation

$$
i_t = \mathbf{Q}[h_t(x_t, v_t) - j_t(\xi_t)] \stackrel{\triangle}{=} \mathfrak{h}_t(z_t, v_t). \qquad (3.19)
$$

• Specializing to the linear Gaussian case with

$$
F_t = \begin{bmatrix} A & 0 \\ L_t C & A - L_t C \end{bmatrix}, n_t = \begin{bmatrix} w_t \\ L_t v_t \end{bmatrix}, H = \begin{bmatrix} C & -C \end{bmatrix},
$$

– state equation

$$
\begin{aligned}
z_{t+1} &= \begin{bmatrix} x_{t+1} \\ \check{x}_{t+1} \end{bmatrix}, \qquad (3.20) \\
&= \begin{bmatrix} A & 0 \\ L_t C & A - L_t C \end{bmatrix} \begin{bmatrix} x_t \\ \check{x}_t \end{bmatrix} + \begin{bmatrix} w_t \\ L_t v_t \end{bmatrix}, \\
&= F_t z_t + n_t, \qquad (3.21)
\end{aligned}
$$

– output equation

$$i_t = \mathbf{Q}\left(\begin{bmatrix} C & -C \end{bmatrix}\begin{bmatrix} x_t \\ \breve{x}_t \end{bmatrix} + v_t\right),$$
$$= \mathbf{Q}\left(Hz_t + v_t\right). \tag{3.22}$$

3.4.1 Open-Loop System Stability Condition

The linear predictive decoder immediately highlights a stability requirement on the source system (3.7) in order that the receiver-side innovations filter also be stable. As the predictive codec is envisaged as part of a feedback control scheme, this imposes a restriction on the class of plants to which such a scheme might be applicable.

Lemma 9. *Consider the transmitter-side linear, time-varying system* (3.20)*, with joint state space* $\mathbb{R}^{2n_x}$*, together with its predictively-coded innovations output signal,* $\{\varepsilon_t\}$ *from* (3.9)*. The subspace*

$$Span\left\{\begin{bmatrix} I_{n_x} \\ I_{n_x} \end{bmatrix}\right\} \subset \mathbb{R}^{2n_x},$$

is unobservable and is associated with the eigenvalues of A.

Proof: Evidently and no matter the value of L_t,

$$\begin{bmatrix} C & -C \end{bmatrix}\begin{bmatrix} I_{n_x} \\ I_{n_x} \end{bmatrix} = 0_{n_y \times n_x},$$
$$\begin{bmatrix} A & 0 \\ L_tC & A - L_tC \end{bmatrix}\begin{bmatrix} I_{n_x} \\ I_{n_x} \end{bmatrix} = \begin{bmatrix} I_{n_x} \\ I_{n_x} \end{bmatrix} \times A.$$

Corollary 10. *Consider the transmitter-side predictive coding system* (3.7)-(3.9),(3.15) *with the*

receiver calculating its state estimate using the innovations filter,

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + L_t i_t. \tag{3.23}$$

If system matrix A has any eigenvalues outside or on the open unit disk and $\check{x}_t - \hat{x}_t$ possesses non-zero component in the direction of this eigenvector for some t, then the error between estimates, $\check{x}_t$ at the transmitter and $\hat{x}_t$ at the receiver, grows unbounded with time.

For stable linear systems at the transmitter, which might consist of the joint stable plant and its stable predictor, the linear analysis of the closed-loop fails when the fixed-range quantizer overflows or saturates. Then, the assumptions of Theorem 8 fail and the quantization error ceases to exhibit the independence properties. Naturally, the Gaussian property of the system noises guarantees both eventual overflow and non-infinitesimal probability of overflow at any time. The probability of saturation of a fixed-quantized signal in these circumstances has been studied using Markov methods by the authors in [32] for both intermittent and quantized data. The time of first overflow is called the escape time there. A feature of that analysis is that for many systems, the escape time can be very large, depending on system parameters including feedback gain K and saturation level ζ.

If A has all eigenvalues in the open unit disk and escape time has yet to occur, then using (3.10), (3.15), (3.23), we have

$$\check{x}_{t+1} - \hat{x}_{t+1} = A(\check{x}_t - \hat{x}_t) - L_t \psi_t,$$

and, letting t grow while vainly betting on no escape,

$$\begin{aligned} \mathrm{E}\left[\hat{x}_t\right] &\to \mathrm{E}\left[\check{x}_t\right] = \mathrm{E}\left[x_t\right], \\ \mathrm{cov}\left[\check{x}_t - \hat{x}_t\right] &\to \sum_{j=0}^{\infty} \left\{ A^j L \Psi L^T A^{j^T} \right\}. \end{aligned}$$

Here, L is the limiting value of the Kalman gain, L_t. These results are independent of the

feedback control law other than central dependence of the escape time itself on K.

These results of Lemma 9 and the discussion following for the underlying linear time-invariant system (3.7) carry over directly to linear time-varying systems by the same argument [33]. For linear systems with nonlinear measurements, one may appeal to Curry [1], who shows that the innovations is independent of the control signal, and Lemma 2 to argue that the unobservability problem persists for these systems. For more general nonlinear systems, it is less clear how instability of the plant might be manifested in the error between transmitter-side and receiver-side estimates. Although, Assumption 2.6 would be needed to analyze the estimate errors locally.

The clear admonition of Corollary 10 is to apply predictive coding solely to the control of stable systems. The reconstruction of the state estimate from the receiver innovations otherwise is unstable. This occurs because there is no output injection of $\hat{x}_t$ into the computation of ε_t and thus i_t. To our knowledge, this was first observed in [34].

We also note that the practically implemented G.722 ADPCM standard [7], in the definition of the adaptive predictor in its Section 3.6, includes specific pole-parameter restrictions to enforce stability of the prediction model at both the transmitter and receiver; it limits the number of poles to two and projects the parameters to ensure stability. Thus, we offer four observations.

(i) Predictive coding does not appear suited to the control of unstable systems. We believe this to be a novel observation and a reflection of the nature of predictive coding itself.

(ii) The practical success of ADPCM indicates that predictive coding has something to offer in control of stable plants.

(iii) The computed feedback control performance, in the example presented in Section 3.6 for a stable linear system close to instability, is significantly improved (reduced by 56%) using predictive coding for a finite bit-rate channel over that in which the output signal itself is quantized.

(iv) The nonlinear example in Section 3.6 also demonstrates significant improvement – in this case with a maximization criterion and by a factor of 15 – of the innovations based approach versus quantization of the output signal.

3.4.2 Bayesian filter

The Bayesian filter uses the sequence of measurements, $\{u_t, i_t\}$, to compute recursively the joint conditional density of the transmitter-side state

$$\pi_t = p\left(z_t | \mathbf{I}^t\right) = p\left(\left[\begin{array}{c} x_t \\ \xi_t \end{array} \right] \middle| \mathbf{I}^t \right).$$

For the general nonlinear system (3.18)-(3.19),

$$\begin{aligned} z_{t+1} &= \mathfrak{f}_t(z_t, u_t, w_t, v_t), \\ i_t &= \mathfrak{h}_t(z_t, v_t), \end{aligned}$$

the Bayesian filter recursion is [35]

$$\begin{aligned} p(z_t|\mathbf{I}^t) &= \frac{p(i_t|z_t, \mathbf{I}^{t-1}) p(z_t|\mathbf{I}^{t-1})}{\int_{z_t} p(i_t|z_t, \mathbf{I}^{t-1}) p(z_t|\mathbf{I}^{t-1})\, dz_t}, \\ &= \frac{p(i_t|z_t, \mathbf{I}^{t-1}) p(z_t|\mathbf{I}^{t-1})}{p(i_t|\mathbf{I}^{t-1})}, \qquad (3.24) \\ p(z_{t+1}|\mathbf{I}^t) &= \int_{z_t} p(z_{t+1}|z_t, \mathbf{I}^t) p(z_t|\mathbf{I}^t)\, dz_t. \qquad (3.25) \end{aligned}$$

We have been careful to include explicitly the conditioning on $\mathbf{I}^t$ in both integrands, since this plays a role in the case of correlated process and measurement noises, as here.

The recursion commences from $\pi_{0|-1} = p(z_0)$ and consists of two parts:

- measurement update (3.24) with $p(i_t|z_t, \mathbf{I}^{t-1})$ derived from the output equation (4.2) via the function $\mathfrak{h}_t(\cdot,\cdot)$ and the density of v_t.

- time update (3.25) with $p(z_{t+1}|z_t, \mathbf{I}^t)$ reflecting $\mathfrak{f}_t(\cdot,\cdot,\cdot,\cdot)$ in (3.18) and the joint densities of w_t, v_t and i_t.

For linear Gaussian systems without quantization, the Bayesian and Kalman filters coincide, although the Kalman filter more efficiently computes just the sufficient statistics of these conditional densities: the mean and covariance.

3.4.3 Reduced-order Bayesian filter

We note that the (full-order) Bayesian filter (3.24)-(3.25) yields the joint density π_t. The marginal densities, $p(x_t|\mathbf{I}^t)$ and $p(\xi_t|\mathbf{I}^t)$, are simply computed from π_t by integration. If, however, only the predictor state density, $p(\xi_t|\mathbf{I}^t)$, is desired, this can more easily be calculated by applying the Bayesian filter to state equation (3.6) with measurement equation (3.12).

$$\begin{aligned} \xi_{t+1} &= g_t(\xi_t, u_t, \varepsilon_t), \quad \xi_0, \\ i_t &= \mathbf{Q}(\varepsilon_t). \end{aligned}$$

In the quantized linear Gaussian case, this corresponds to using (3.10) and (3.12) to compute the density $p(\check{x}_t|\mathbf{I}^t)$ without the attendant calculation of $p(x_t|\mathbf{I}^t)$. This results in a reduced-order Bayesian filter which yields solely the conditional density of $\check{x}_t$. Such receiver-side reconstruction of the predictor state is the mainstay of predictive coding in signal processing [6]. Such ideas underpin some approaches to quantized innovations Kalman and Bayesian filtering [10, 11, 12] and Delta Modulation.

3.4.4 Computational issues

The Bayesian filter is numerically demanding. Notably, the integration in time update (3.25) presents a challenge to computation, since it involves performing a $2n_x$-dimensional integral at each sample point in a $2n_x$-dimensional space, yielding an operation count of $\mathscr{O}\left(16n_x^4\right)$. By contrast, the measurement update (3.24) is relatively benign at $\mathscr{O}(4n_x^2)$. Increasing the number of sample points per dimension rapidly causes problems. This is exacerbated by densities in

x_t and $\check{x}_t$ being poorly conditioned, such as can occur with singular densities for $\check{x}_t$ and with very fine quantization. No special numerical 'tricks' were applied in the computations in this paper. The Particle filter may be applied to implement approximately the Bayesian filter using resampling ideas to manage calculations. This comes with its own set of problems, issues and fixes [36]. In our examples, we compute the Bayesian filter on a fixed grid rather than by particles.

The distinction between computation of $p(x_t|\mathbf{I}^t)$ and $p(\check{x}_t|\mathbf{I}^t)$ rests solely with the willingness to devote resources to computation at the receiver. They both operate on the same data. In a control setting, this is also connected to the admissibility of accepting greater computational delay at the receiver, which itself might preclude any advantage versus the delay in accepting a prediction-based control signal.

It is certainly worth remarking that the Bayesian filter calculations, notably central recursion (3.25), lend themselves to highly parallelized implementation, which suggests using GPUs or other processor architectures to achieve speedup [37].

3.4.5 Density properties

We have the following general results for the nonlinear and linear cases.

Theorem 11. *If the conditional mean state estimate is computed at the transmitter, so that*

$$\check{x}_t = E(x_t|\mathbf{E}^{t-1}),$$

then the two receiver-side conditional means coincide. That is,

$$E(x_t|\mathbf{I}^{t-1}) = E(\check{x}_t|\mathbf{I}^{t-1}). \tag{3.26}$$

Proof: Lemma 7 shows that $\sigma(\mathbf{I}^{t-1}) = \mathscr{I}_{t-1} \subseteq \mathscr{E}_{t-1} = \sigma(\mathbf{E}^{t-1})$. The *smoothing property*

of conditional expectation [38] then establishes that

$$\mathrm{E}[\check{x}_t|\mathbf{I}^{t-1}] = \mathrm{E}\left[\mathrm{E}[x_t|\mathbf{E}^{t-1}]|\mathbf{I}^{t-1}\right] = \mathrm{E}[x_t|\mathbf{I}^{t-1}].$$

Theorem 12. *In the general nonlinear case, if the innovations sequence, $\{\varepsilon_t\}$, is white, then*

$$p(\check{x}_t|\mathbf{I}^t) = p(\check{x}_t|\mathbf{I}^{t-1}).$$

Proof: The whiteness property of $\{\varepsilon_t\}$ implies that the received signal, $\{i_t\}$, also is white and, since $\check{x}_t$ is computed causally from $\mathbf{E}^{t-1}$, that $\check{x}_t$ is independent from ε_t and, therefore, i_t.

• Theorem 11 states that, should the objective be to calculate the conditional mean of the plant state at the receiver, then one might use the reduced-order Bayesian filter to achieve this.

• Theorem 12 establishes that in the case where the prediction errors are white, the conditional $\check{x}_t$ density at the receiver (and transmitter) will update only at the time-update stage.

• We appreciate that, while the transmitter side recursion (3.3) is driven by the innovations, ε_t, derived directly from x_t, the receiver side Bayesian filter driven by the quantized innovations requires stability of $g_t(\cdot,\cdot,\cdot)$ as in Assumption 2.6 in order that its predictor state estimate not diverge too greatly from that at the transmitter. This, underlying predictor stability requirement is inherent in all works in this field and reflects the estimate convergence condition for two state estimators both driven by the innovations of one of the estimators.

3.5 Controller

The sequence of quantized innovations, $\{i_t\}$, arrives at the receiver and is used to generate the feedback control signal, u_t, as depicted Figure 3.2. The Bayesian filter is applied to the received sequence to yield conditional densities $p(x_t|\mathbf{I}^{t-1})$ and $p(\check{x}_t|\mathbf{I}^t)$. We have the following result from stochastic optimal control.

Theorem 13 (Kumar & Varaiya [39], Bertsekas [40])**.** *For any choice of optimization criterion*

admitting a bounded value function, the optimal causal output feedback control for system (3.18)-(3.19) *is*

$$u_t^{opt} = \mathfrak{k}_t(\pi_t),$$

where $\pi_t = p(z_t|\mathbf{I}^t)$ *and feedback policy* $\mathfrak{k}_t(\pi_t)$ *is found by solving the stochastic dynamic programming equation based on the associated objective function.*

The result follows from the Markovian property of (3.18) and involves two computationally challenging aspects; the Bayesian filter for π_t and the solution of the stochastic dynamic programming equation. Inherently, this latter piece is the harder and requires duality of the controller. For our predictive coding setup, since the plant state x_t is not dependent on ξ_t, we can say more.

Corollary 14. *For system* (4.1)-(4.2)*, with objective function dependent solely on future* $\{x_t, u_t\}$*, the optimal causal output feedback control solution is*

$$u_t^{opt} = k_t\left(p(x_t|\mathbf{I}^t)\right).$$

That is to say, the predictor state, ξ_t, and its conditional density are not explicitly part of the optimal control solution.

The control signal computation is based on the conditional x_t density from the Bayesian filter. Our central aim is to describe the Bayesian filter for estimating the joint state,

$$z_t = \begin{bmatrix} x_t \\ \breve{x}_t \end{bmatrix},$$

representing the predictively coded transmitter side.

Theorem 15 (Curry [1], Section 5.4, pp. 75-78, Appendix D, pp. 114-116)**.** *For the linear state*

system (3.7) *with: memoryless nonlinear measurement*

$$y_t = \varphi_t(x_t, v_t),$$

independent, white but not necessarily Gaussian noise processes $\{w_t\}$ *and* $\{v_t\}$*, and quadratic objective function*

$$J_t = E\left(\sum_{k=t}^{N+1} x_k^T Q_k x_k + u_k^T R_k u_k \middle| \mathbf{Y}^t, \mathbf{U}^{t-1}, \pi_{0|-1} \right), \tag{3.27}$$

the optimal output feedback control is given by

$$u_t^\star = K_t\, E(x_t | \mathbf{Y}^t, \mathbf{U}^{t-1}, \pi_{0|-1}),$$

where, K_t *is the LQ optimal feedback gain computed from the control Riccati equation.*

Corollary 16. *For linear system* (3.7) *with quantized innovations measurement* (3.12)*, quadratic objective function* (4.3)*, and one time-sample delay in the controller, the optimal output feedback control is given by*

$$u_t^\star = K_t\, E(\check{x}_t | \mathbf{I}^{t-1}, \mathbf{U}^{t-1}, \pi_{0|-1}). \tag{3.28}$$

Proof: Mita [41] establishes the optimality of the LQ-optimal feedback gain with the predictive state estimate. This translates directly to Curry's result. For linear systems, the innovations sequence is white and one may then appeal to Theorem 11 to establish that $\mathrm{E}(x_t | \mathbf{I}^{t-1}, \mathbf{U}^{t-1}, \pi_{0|-1}) = \mathrm{E}(\check{x}_t | \mathbf{I}^{t-1}, \mathbf{U}^{t-1}, \pi_{0|-1})$ and the result follows.

• Theorem 17 shows that, despite the nonlinear measurements, the optimal LQ control is to feed back the filtered conditional mean of the state. This would suggest using the receiver to compute this quantity and then to calculate the control.

• Appealing to the results of [41], we see that, for a single delay controller, the same calculation

holds but with the predicted state estimate at the receiver, which is simpler to compute.

• Corollary 16 uses Theorem 11 to replace the conditional mean of x_t by that of $\check{x}_t$, which incurs substantially fewer computations for its estimation.

• It is worth noting that the filtered conditional mean of x_t is different from its predicted conditional mean, even though Theorem 12 shows that they coincide for $\check{x}_t$ when the innovations is white, as in the linear case.

• This theorem and corollary are specialized to linear systems with quadratic criteria. We shall see shortly an example, where the optimal controller depends on the complete density of x_t and not just on $\check{x}_t$. In this case, the feedback controller based on $p(x_t|\mathbf{I}^{t-1})$ outperforms that based on $p(\check{x}_t|\mathbf{I}^{t-1})$.

3.6 Quantized Linear Innovations Filtering

We now specialize the development to the case of quantized linear Gaussian innovations, the Bayesian filter for which is derived in the Appendix. In this section, we do not use a subtractive dithered quantizer and compute the full Bayesian filter. In the following section, we apply the subtractive dithered quantizer and avail ourselves of the whiteness and uniform density of the quantization noise.

The joint state z_t is defined in (3.20) and evolves according to the linear dynamics in (3.21), which defines system matrix F_t. The quantized innovations signal, i_t, is described by (3.22), which defines output matrix H. The Bayesian filter generates: the conditional density, $p(z_t|\mathbf{I}^t)$, of this $2n_x$-dimensional state, the marginal densities of which yield $p(x_t|\mathbf{I}^t)$, to be used for the optimal controller; and, $p(\check{x}_t|\mathbf{I}^t) = p(\check{x}_t|\mathbf{I}^{t-1})$ according to Theorem 12.

By the same token, we also consider the n_x-dimensional Bayesian filter for the innovations representation of the transmitter-side state estimator,

$$\begin{aligned} \check{x}_{t+1} &= A\check{x}_t + Bu_t + L_t\varepsilon_t, \\ i_t &= \mathbf{Q}(\varepsilon_t), \end{aligned} \tag{3.29}$$

to construct directly $p(\check{x}_t|\mathbf{I}^t)$, without the attendant complication of producing $p(x_t|\mathbf{I}^t)$, nor indeed of performing the measurement update step. The conditional density $p(\check{x}_t|\mathbf{I}^t)$ is identical whether produced via the full-order Bayesian filter or it reduced-order counterpart.

Example system

We consider a scalar example quantized innovations system with: values $A = 0.99$, $B = 1$, $C = 1$, $Q = \text{cov}(w_t) = 0.1$, $R = \text{cov}(v_t) = 0.1$; Kalman filter initialization $\hat{x}_{0|-1} = 0$, $\Sigma_{0|-1} = 1.3$. Depending on the signal, ε_t or y_t, being quantized, the corresponding steady-state standard deviation, σ_ε or σ_y, is computed and the 3-bit/8-level, linear, symmetric, midrise quantizer is used with saturation value at $\zeta_\varepsilon = 5\sigma_\varepsilon$ or $\zeta_y = 5\sigma_y$ respectively, where σ refers to the stationary variance.

The quantized innovations Bayesian filter, the quantized output Bayesian filter, and the unquantized Kalman filter were computed for a number of steps. The resulting predicted and filtered densities are displayed in Figures 3.3 and 3.4. The densities were propagated at 71 sample points in the range [-2, 2].

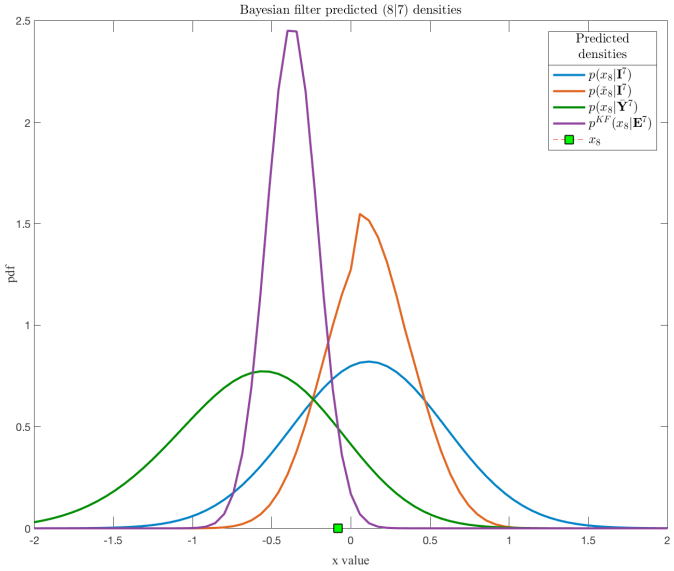

Figure 3.3. Predicted $(8|7)$ density functions for: quantized innovations Bayesian filter $p(x_8|\mathbf{I}^7)$ and $p(\check{x}_8|\mathbf{I}^7)$, quantized output Bayesian filter $p(x_8|\bar{\mathbf{Y}}^7)$, transmitter-side Kalman predictor $p^{KF}(x_8|\mathbf{E}^7)$. Actual plant state x_8 depicted by a green square.

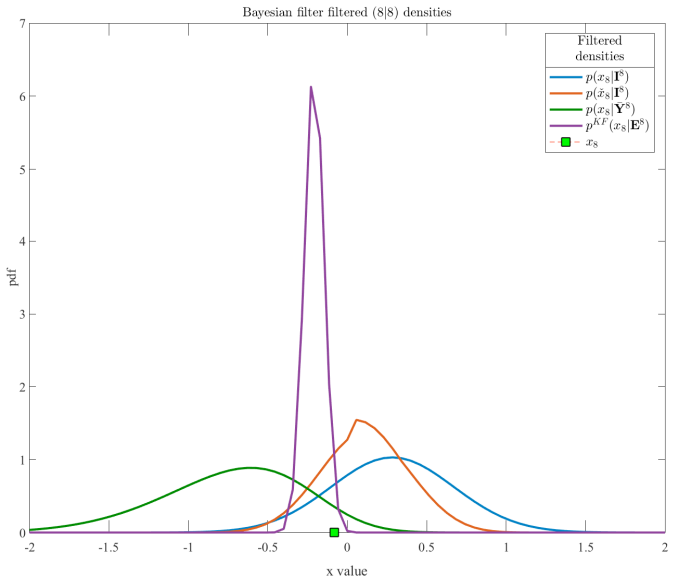

Figure 3.4. Filtered $(8|8)$ density functions for: quantized innovations Bayesian filter $p(x_8|\mathbf{I}^8)$ and $p(\check{x}_8|\mathbf{I}^8)$, quantized output Bayesian filter $p(x_8|\bar{\mathbf{Y}}^8)$, transmitter-side Kalman predictor $p^{KF}(x_8|\mathbf{E}^8)$. Actual plant state x_8 depicted by a green square. Note change of vertical scale versus Figure 3.3.

We offer the following observations.

• The predicted densities $p(x_8|\mathbf{I}^7)$ and $p(\check{x}_8|\mathbf{I}^7)$ are different, although their mean values are the same, as guaranteed by Theorem 11.

• The filtered densities $p(x_8|\mathbf{I}^8)$ and $p(\check{x}_8|\mathbf{I}^8)$ are different, as are their mean values.

• The filtered density $p(\check{x}_8|\mathbf{I}^8)$ is identical to the predicted density $p(\check{x}_8|\mathbf{I}^7)$, as guaranteed by Theorem 12.

• The conditional densities of x_t based on quantized innovations are different from those based on quantized output y_t.

3.7 Linear Innovations with Dithered Quantizer

We now replace the standard quantizer by a subtractive dithered quantizer, as described in Theorem 8, where now the quantization noise is assumed white and uniformly distributed with zero mean and covariance $\frac{\zeta^2}{3\times 2^{2b}}$. The Kalman filter provides optimal second-order estimation in this case.

We have the above linear transmitter systems (3.18) and (3.10) with transmitted data

$$i_t = \varepsilon_t + \psi_t, \tag{3.30}$$

$$= Hz_t + v_t + \psi_t. \tag{3.31}$$

Immediately, one has the receiver-side Kalman filter recursion from (3.21)-(3.30) with usual accommodation of correlated process and measurement noises. Denote

$$\mathscr{F}_t = F_t \bar{P}_t H^T + \begin{bmatrix} 0 \\ L_t R \end{bmatrix}.$$

Then the recursion for the receiver's conditional mean and conditional covariance is:

$$\mu_{t+1} = F_t \mu_t + \mathscr{F}_t (H \bar{P}_t H^T + R + \Psi)^{-1} i_t, \tag{3.32}$$

$$\begin{aligned} \bar{P}_{t+1} = {} & F_t \bar{P}_t F_t^T - \mathscr{F}_t (H \bar{P}_t H^T + R + \Psi)^{-1} \mathscr{F}_t^T \\ & + \begin{bmatrix} Q & 0 \\ 0 & L_t R L_t^T \end{bmatrix} \end{aligned} \tag{3.33}$$

with initial condition

$$\mu_0 = \begin{bmatrix} \hat{x}_{0|-1} \\ \hat{x}_{0|-1} \end{bmatrix}, \quad \bar{P}_0 = \begin{bmatrix} P_{0|-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

The conditional joint density at the receiver has

$$\mathrm{E}(z_{t+1} | \mathbf{I}^t) = \mu_t,$$

$$\mathrm{cov}(z_{t+1} | \mathbf{I}^t) = \bar{P}_t.$$

Observations

• The first two conditional moments of $p(\check{x}_t|\mathbf{I}^{t-1})$ are computed directly from signal model (3.10) and (3.30).

$$\begin{aligned}\hat{\check{x}}_{t+1} &= \mathrm{E}[\check{x}_{t+1}|\mathbf{I}^t, \pi_{0|-1}], \\ &= A\hat{\check{x}}_t + Bu_t + L_t i_t, \qquad (3.34)\\ M_{t+1} &= \mathrm{cov}(\check{x}_{t+1}), \\ &= AM_tA^T + L_t\Sigma_{t|t-1}[I-(\Sigma_{t|t-1}+\Psi)^{-1}\Sigma_{t|t-1}]L_t^T. \qquad (3.35)\end{aligned}$$

The resultant conditional density is unique no matter the method of computation.

• The detailed recursion (3.34)-(3.35) for $\hat{\check{x}}_t$ shows that the estimate only adjusts at the time-update step of the Kalman filter, since $p(\check{x}_t|\mathbf{I}^t) = p(\check{x}_t|\mathbf{I}^{t-1})$, which in turn is due to the independence of $\check{x}_t$ and i_t from (3.34). This is a manifestation of Theorem 12.

• The two dimension-n_x components of the conditional mean are equal per Theorem 11.

$$\mathrm{E}(x_t|\mathbf{I}^{t-1}) = \mathrm{E}(\check{x}_t|\mathbf{I}^{t-1}).$$

• While the conditional means of x_t and $\check{x}_t$ at the receiver are identical in this case, their covariances, and thus their complete conditional densities are different.

• For the case with zero channel or quantization noise, ψ_t,

$$\bar{P}_t = \mathrm{blockdiag}(\Sigma_{t|t-1}\ \ 0).$$

In this case, $\check{x}_t$ is reconstructed perfectly at the receiver, since it is a deterministic function of the innovations.

• The stability of matrix A is required for the convergence of conditional mean M_t in (3.35). This follows Corollary 10.

• We have deliberately ignored the condition $\varepsilon_t \in [-\zeta, \zeta]$ from Theorem 8, required to ensure

that ψ_t be white. The saturation levels on the quantizer are presumed chosen to enforce this with high probability.

• Additive white channel noise could be included into the analysis in a fashion identical to the quantization noise, subject to the saturation condition.

3.8 Comparative optimal control examples

We use the scalar example system from Section 3.6 above and consider a sequence of control problems applied to the subtractive dithered quantized system. The aim is to identify circumstances where the control benefits accrue with the availability of the filtered state density.

3.8.1 LQG control with dithered quantizer

For the example system presented earlier and LQ cost function

$$J = \lim_{N\to\infty} \mathrm{E}\left\{ \frac{1}{N} \sum_{t=0}^{N-1} x_t^T Q_c x_t + u_t^T R_c u_t \right\},$$

with $Q_c = 5$ and $R_c = 0.7$, the performance of three controllers was computed using the LQ-optimal feedback gain and the various conditional mean state estimates.

I. filtered state estimate from quantized innovations

$u_t = -K\mathrm{E}(x_t|\mathbf{I}^t)$,

[This is the optimal control by Theorem 17.]

II. predicted state estimate from quantized innovations

$u_t = -K\mathrm{E}(x_t|\mathbf{I}^{t-1}) = -K\mathrm{E}(\check{x}_t|\mathbf{I}^{t-1})$,

III. filtered state estimate from quantized outputs

$u_t = -K\mathrm{E}(x_t|\bar{\mathbf{Y}}^t)$.

The achieved LQG costs are given in this table.

Control law	value J
$u_t = -K\mathrm{E}(x_t\|\mathbf{I}^t)$	0.9311
$u_t = -K\mathrm{E}(\check{x}_t\|\mathbf{I}^{t-1})$	1.4126
$u_t = -K\mathrm{E}(x_t\|\bar{\mathbf{Y}}^t)$,	1.6712

These quantifications indicate the following.

• The optimal control relies on the use of the filtered density and there is a substantial performance penalty to using the predictive density for this example. This needs to be balanced against the computational cost of operating the full-order Bayesian filter at the receiver.

• There is, for this example, a substantial performance benefit accruing the efficient use of the communications channel through the transmission of quantized innovations signals versus quantized outputs. Again, this comes at a complexity cost in computation at the receiver. But it shows that a control improvement can be realized via careful signal coding. This is generally well understood [19, 20] but is quantified by the example here.

3.8.2 Non-LQ optimal control

For the same system, define the performance function

$$\eta_t = \begin{cases} x_t, & \text{if } x_t < 1, \\ 0, & \text{else.} \end{cases}$$

The one-step-ahead control objective is

$$u_t = \mathrm{argmax}_{u_t} \mathrm{E}\left[\eta_{t+1}|\mathbf{I}^t\right].$$

The solution for the optimal control, given by (3.42) in the Appendix, is

$$u^{\mathrm{opt}} = 1 - \mathrm{E}(x_{t+1}^{\mathrm{NF}}|\mathbf{I}^t) - \mathrm{xsolv},$$

where xsolv is the solution of (3.43), an algebraic equation involving solely the Bayesian filter predicted density function $p^{\mathrm{NF}}(x_{t+1}^{\mathrm{NF}}|\mathbf{I}^t)$ for the unforced, $u_t = 0$, state. That is, the optimal control depends on the entire density of the state and not just upon its first moment. The formula (3.45) for the optimal cost in this case indicates dependence on the covariance.

We present three examples of optimal control of η_t based on the densities: $p(x_{t+1}|\mathbf{I}^t)$, $p(\check{x}_{t+1}|\mathbf{I}^t)$ and $p(x_t|\bar{\mathbf{Y}}^t)$. As seen from Figure 3.3, these densities differ and each is associated with a different value of the control parameters xsolv. Accordingly, their control performances differ, even though their conditional means might coincide.

Density	xsolv	$\mathrm{E}[\eta_{t+1}\|\mathbf{I}^t]$
$p(x_{t+1}\|\mathbf{I}^t)$	-0.0673	0.1623
$p(\check{x}_{t+1}\|\mathbf{I}^t)$	0.0587	0.1587
$p(x_t\|\bar{\mathbf{Y}}^t)$	-2.8118	0.0108

This reinforces the control performance value of the use of the filtered quantized innovations state density. The xsolv value computed from the $\check{x}$ density is inappropriate, leading to diminished performance. For the quantized output density, the increased variance in the density due to inefficient coding degrades control performance.

3.9 Conclusion & extensions

We have explored the application of the Bayesian filter for control based on predictively coded signals. The predictive coding brings efficiency in the use of the channel bits, which leads to improved state estimation at the receiver and, in turn, to a more accurate state density for control calculation. We have paid particular attention to the generation of the filtered state conditional density, the *information state*, at the controller and identified the inherent performance difference from the predicted state conditional density.

In addition to new theoretical results concerning the state estimation task with predictive coding, the demonstration of computed examples illustrates the feasible but high computational cost of these methods. We analyzed the control problem with dithered quantization which permits

precise evaluation of control performance in several cases. Extensions of the computational examples are possible.

- Computation with fully nonlinear and time-varying state and measurement equations, as illustrated in Figure 3.2, requires some finesse in the following manner.
  - The transmitter-side predictor yielding ξ_t and $\check{y}_t$ needs to be based itself on a nonlinear filter, perhaps even a Bayesian filter.
  - The conditional densities $p(i_t|z_t)$ in (3.24) need to incorporate the nonlinearity $h_t(\cdot,\cdot,\cdot)$ in an appropriate fashion in addition to the inclusion of the quantizer.
  - The conditional densities $p(z_{t+1}|z_t)$ in (3.25) need to include the nonlinearity $f_t(\cdot,\cdot,\cdot)$.
  - The innovations sequence no longer need be white. Even in the linear non-Gaussian case, it is uncorrelated but not necessarily white.

  These are standard issues with the application of the Bayesian filter.

- Incorporation of further channel defects such as dropped packets, additive noise, delays are simple extensions of the Bayesian filter. We have already commented on additive channel noise above.

- Practical issues arise when implementing the Bayesian filter. Here, because we have chosen a stationary problem, we have been able to compute the conditional densities on a static grid in the z_t-space. More generally, the Bayesian filter is realized via the Particle filter [36]. This requires some skill.

The authors are keen to acknowledge the technically sound and very helpful comments and guidance from the reviewers.

3.10 Appendix

3.10.1 Proof of Theorem 8

Theorem QTSD of [3], Section 19.8, pp. 506-512, states that, for a linear quantizer with quantization interval q, provided the characteristic function, $\Phi_d(\cdot)$, of the subtractive dither signal satisfies

$$\Phi_d\left(l\frac{2\pi}{q}\right) = 0, \text{ for } l = \pm 1, \pm 2, \ldots, \tag{3.36}$$

and

$$\varepsilon + d \in [-\zeta, \zeta], \tag{3.37}$$

then the quantization error, $\varepsilon - \mathscr{Q}(\varepsilon)$, will be independent of the input signal, ε, and uniformly distributed $\mathscr{U}(-q/2, q/2)$.

For the linear quantizer of range 2ζ and 2^b levels, $q = \zeta/2^{b-1}$. The characteristic function of a $\mathscr{U}[-a,a]$ density is $\Phi_{d,\text{unif}}(\omega) = \operatorname{sinc} a\omega$. Taking $a = q/2$, $\Phi_{d,\text{unif}}(\omega) = \operatorname{sinc} q\omega/2$, which satisfies (3.36) above.

The pdf of the sum of two independent $\mathscr{U}(-q/2, q/2)$ random variables is the convolution of the uniform pdfs and is triangularly distributed $\text{tr}(-q, q)$. By the properties of the Fourier transform, $\Phi_{d,\text{tr}}(\omega) = \Phi^2_{d,\text{unif}}(\omega)$ and (3.36) is satisfied.

We note in passing that Theorem QTSD does not explicitly state the saturation condition (3.37) on the additively dithered signal. Without it, the theorem fails.

3.10.2 Derivation of the Bayesian filters for quantized linear systems

In Section 3.8, we explore the optimal control performance of three candidate approaches to state conditional density reconstruction at the receiver.

1. Full $2n_x^{\text{th}}$-order quantized innovations Bayesian filter, reconstruction of the conditional

state density $p(x_t|\mathbf{I}^t)$, and computation of the optimal control using this density.

2. Simplified n_x^{th}-order quantized innovations Bayesian filter, reconstruction of the conditional state estimate density $p(\check{x}_t|\mathbf{I}^t)$, and computation of the optimal control using this density.

3. The n_x^{th}-order Bayesian filter operating directly on the quantized output signal, $\phi_t = \mathbf{Q}(y_t)$, reconstruction of the conditional density $p(x_t|\blacksquare^t)$, and computation of the optimal control using this density.

We now present the detailed Bayesian filter for each case.

3.10.3 Bayesian filter for quantized innovations

Measurement update

We begin the Bayesian filter recursion from the predicted density $p(z_t|\mathbf{I}^{t-1})$ with the current measurement i_t in hand. This i_t corresponds to $\varepsilon_t \in (\varepsilon_{\text{lower}_t}, \varepsilon_{\text{upper}_t}]$. Then, from (3.21)-(3.22),

$$\begin{aligned} p(i_t|z_t,\mathbf{I}^{t-1}) &= p(i_t|z_t), \\ &= \int_{\varepsilon_{\text{lower}_t}}^{\varepsilon_{\text{upper}_t}} p(v_t = \varepsilon_t - Cx_t + C\check{x}_t)\, dv_t, \\ &= \texttt{mvncdf}(\varepsilon_{\text{lower}_t}, \varepsilon_{\text{upper}_t}, Hz_t, R). \end{aligned} \tag{3.38}$$

The Matlab function `mvncdf` computes the multivariate normal cumulative distribution function between lower and upper limits with given mean and covariance. This is then used in (3.24) to yield the filtered joint conditional density $p(z_t|\mathbf{I}^t)$.

Time update

For the time update step (3.25), the system equations (3.18) and (3.11) yield

$$\begin{aligned} w_t &= x_{t+1} - Ax_t - Bu_t, \\ L_t v_t &= \check{x}_{t+1} - L_t C x_t - (A - L_t C)\check{x}_t - Bu_t. \end{aligned}$$

Whence, the conditional density

$$\begin{aligned} p(z_{t+1}|z_t, \mathbf{I}^t) &= \frac{p(z_{t+1}, i_t|z_t, \mathbf{I}^{t-1})}{p(i_t|z_t, \mathbf{I}^{t-1})}, \\ &= \frac{p(z_{t+1}, i_t|z_t, \mathbf{I}^{t-1})}{p(i_t|z_t)}, \end{aligned} \tag{3.39}$$

since the innovations and quantized innovations, $i_t = \mathbf{Q}(\varepsilon_t)$, are white. Denominator $p(i_t|z_t)$ is given by (3.38). The numerator comprises three terms

$$p(z_{t+1}, i_t|z_t, \mathbf{I}^{t-1}) = W \times V \times T, \tag{3.40}$$

with

$$\begin{aligned} W &= p(w_t = x_{t+1} - Ax_t - Bu_t), \\ &= \texttt{mvnpdf}(x_{t+1} - Ax_t - Bu_t, 0, Q), \\ V &= p(L_t v_t = \check{x}_{t+1} - L_t C x_t - (A - L_t C)\check{x}_t - Bu_t), \\ &= \texttt{mvnpdf}(\check{x}_{t+1} - L_t C x_t - (A - L_t C)\check{x}_t - Bu_t, 0, L_t R L_t^T), \\ T &= \mathbf{1}(Cx_t - C\check{x}_t + v_t \in (\varepsilon_{\text{lower}_t}, \varepsilon_{\text{upper}_t}]). \end{aligned}$$

Here, Matlab function `mvnpdf` is the multivariate normal probability density function and $\mathbf{1}(\cdot)$ is the set indicator function. Relations (3.38) and (3.40) comprise the parts of (3.39) of the time update step (3.25) of the Bayesian filter.

3.10.4 Bayesian filter for state-estimate density calculation

In place of the $2n_x$-dimension Bayesian filter (3.24)-(3.25) using relations (3.38)-(3.40), we may appeal to (3.10) and (3.12) as the basis of an n_x-dimensional Bayesian filter for $p(\check{x}_t|\mathbf{I}^t)$. The system equations are

$$\begin{aligned}\check{x}_{t+1} &= A\check{x}_t + Bu_t + L_t\varepsilon_t,\\ i_t &= \mathbf{Q}(\varepsilon_t),\\ \check{x}_0 &= \hat{x}_{0|-1}.\end{aligned}$$

The driving noise process, $\{\varepsilon_t\}$, is white and Gaussian with the following density

$$\varepsilon_t \sim \mathcal{N}(0, C\Sigma_{t|t-1}C + R).$$

Further, this whiteness together with the $\check{x}_t$ update (3.11) ensures that $\check{x}_t$ is independent from ε_t. Thus,

$$\begin{aligned}p(i_t|\check{x}_t) &= p(i_t)\\ &= \texttt{mvncdf}(\varepsilon_{\text{lower}_t}, \varepsilon_{\text{upper}_t}, 0, C\Sigma_{t|t-1}C^T + R).\end{aligned} \tag{3.41}$$

Also, similarly to earlier,

$$\begin{aligned}p(\check{x}_{t+1}|\check{x}_t, \mathbf{I}^t) &= \frac{p(\check{x}_{t+1}, i_t|\check{x}_t, \mathbf{I}^{t-1})}{p(i_t|\check{x}_t, \mathbf{I}^{t-1})},\\ &= \frac{p(\check{x}_{t+1}, i_t|\check{x}_t, \mathbf{I}^{t-1})}{p(i_t)}.\end{aligned}$$

3.10.5 Bayesian filter for quantized outputs

This now proceeds directly from (4.1)-(4.2). Central quantities,

$$p(\bar{y}_t|x_t,\bar{\mathbf{Y}}^{t-1}) = p(\bar{y}_t|x_t),$$
$$p(x_{t+1}|x_t,\bar{\mathbf{Y}}^{t}) = p(x_{t+1}|x_t),$$

are fully described by, respectively: $h_t(\cdot,\cdot)$ and the density of v_t; and $f_t(\cdot,\cdot,\cdot)$ and the density of w_t.

For the linear systems case,

$$p(\bar{y}_t|x_t) = \texttt{mvncdf}(\varepsilon_{\text{lower}_t},\varepsilon_{\text{upper}_t},Hx_t,R),$$
$$p(x_{t+1}|x_t) = \texttt{mvnpdf}(x_{t+1}-Ax_t-Bu_t,0,Q).$$

These expressions extend simply for nonlinear system equations involving solely additive noises.

3.10.6 Optimal control and value $\mathbf{E}[\eta_{t+1}|\mathbf{I}^t]$ for system (3.7)

The predicted state density generated by the Bayesian filter is $p^{\text{NF}}(x_{t+1}^{\text{NF}}|\mathbf{I}^t)$, the unforced, i.e. $u_t = 0$, state since u_t has yet to be determined. Eventually, $x_{t+1} = x_{t+1}^{\text{NF}} + u_t$. Denote conditional mean $\mu_{t+1} = \text{E}(x_{t+1}^{\text{NF}}|\mathbf{I}^t)$ and define the centered unforced state $x_{t+1}^{\text{c}} = x_{t+1}^{\text{NF}} - \mu_{t+1}$. Then the forced state is decscribed by $x_{t+1} = x_{t+1}^{\text{c}} + \mu_{t+1} + u_t$. Thus,

$$\begin{aligned}
E[\eta_{t+1}|\mathbf{I}^t] &= \int_{-\infty}^{1} x_{t+1}\, p(x_{t+1}|\mathbf{I}^t)\, dx_{t+1},\\
&= \int_{-\infty}^{1-\mu_{t+1}-u_t} (x_{t+1}^{\text{c}} + \mu_{t+1} + u_t)\, p^{\text{c}}(x_{t+1}^{\text{c}}|\mathbf{I}^t)\, dx_{t+1}^{\text{c}}\\
&= \int_{-\infty}^{1-\mu_{t+1}-u_t} x_{t+1}^{\text{c}}\, p^{\text{c}}(x_{t+1}^{\text{c}}|\mathbf{I}^t)\, dx_{t+1}^{\text{c}}\\
&\quad + (\mu_{t+1}+u_t) \int_{-\infty}^{1-\mu_{t+1}-u_t} p^{\text{c}}(x_{t+1}^{\text{c}}|\mathbf{I}^t)\, dx_{t+1}^{\text{c}}
\end{aligned}$$

Differentiating with respect to u_t,

$$\begin{aligned}\frac{dE[\eta_{t+1}|\mathbf{I}^t]}{du_t} &= (\mu_{t+1}+u_t-1)\,p^{\mathrm{c}}(1-\mu_{t+1}-u_t|\mathbf{I}^t)\\ &\quad-(\mu_{t+1}+u_t)\,p^{\mathrm{c}}(1-\mu_{t+1}-u_t|\mathbf{I}^t)\\ &\quad+\int_{-\infty}^{1-\mu_{t+1}-u_t} p^{\mathrm{c}}(x^{\mathrm{c}}_{t+1}|\mathbf{I}^t)\,dx^{\mathrm{c}}_{t+1}.\end{aligned}$$

Setting this derivative to zero yields the optimal control

$$\begin{aligned}u^{\mathrm{opt}} &= 1-\mu_{t+1}-\mathrm{xsolv},\\ &= 1-\mathrm{E}(x^{\mathrm{NF}}_{t+1}|\mathbf{I}^t)-\mathrm{xsolv},\end{aligned} \tag{3.42}$$

Where xsolv satisfies

$$p^{\mathrm{NF}}(\mathrm{xsolv}|I_{k-1}) = \int_{-\infty}^{\mathrm{xsolv}} p^{\mathrm{NF}}(x^{\mathrm{NF}}_{t+1}|\mathbf{I}^t)dx^{\mathrm{NF}}_{t+1}. \tag{3.43}$$

Recall that $p^{\mathrm{NF}}(x^{\mathrm{NF}}_{t+1}|\mathbf{I}^t)$ is the state predicted density produced by the Bayesian filter. Thus xsolv is the point where the probability density function crosses the cumulative distribution function for x^{NF}_{t+1}.

The optimal value function or performance is given by

$$\begin{aligned}\mathrm{E}(\eta_{t+1}|\mathbf{I}^t) &= \int_{-\infty}^{\mathrm{xsolv}} x^{\mathrm{c}}_{t+1}\,p^{\mathrm{c}}(x^{\mathrm{c}}_{t+1}|\mathbf{I}^t)\,dx^{\mathrm{c}}_{t+1}\\ &\quad+(1-\mathrm{xsolv})\int_{-\infty}^{\mathrm{xsolv}} p^{\mathrm{c}}(x^{\mathrm{c}}_{t+1}|\mathbf{I}^t)\,dx^{\mathrm{c}}_{t+1}.\end{aligned} \tag{3.44}$$

If the Bayesian filter predicted state density, $p^{\mathrm{NF}}(x^{\mathrm{NF}}_{t+1}|\mathbf{I}^t)$, is Gaussian $\mathcal{N}(\mu_{t+1},\sigma^2)$ then

$$\begin{aligned}\mathrm{E}(\eta_{t+1}|\mathbf{I}^t) &= -\frac{\sigma}{\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(\mathrm{xsolv})^2\right]\\ &\quad+(1-\mathrm{xsolv})\,\mathrm{normcdf}(\mathrm{xsolv},0,\sigma).\end{aligned} \tag{3.45}$$

ACKNOWLEDGEMENTS

I would like to acknowledge Chapter 3 is from the paper,“Predictive coding and control”, IEEE Transaction On Control of Network Systems, , Chun-Chia Huang, Behrooz Amini and Robert R. Bitmead 2018. Chapter 4 is from the paper,“LQG Control Performance with Low Bitrate Periodic Coding,” submitted to IEEE Transaction on Control of Network Systems, Behrooz Amini, Robert R. Bitmead 2020.

Chapter 4

LQG Control Performance with Low Bitrate Periodic Coding

Abstract

Specific low-bitrate coding strategies are examined through their effect on LQ control performance. By limiting the subject to these methods, we are able to identify principles underlying coding for control; a subject of significant recent interest but few tangible results. In particular, we consider coding the quantized output signal deploying period-two codes of differing delay versus accuracy tradeoff. The quantification of coding performance is via the LQ control cost. The feedback control system comprises the coder-decoder in the path between the output and the state estimator, which is followed by linear state-variable feedback, as is optimal in this case. The quantizer is treated as the functional composition of an infinitely-long linear staircase function and a saturation. This permits the analysis to subdivide into estimator computations, seemingly independent of the performance criterion, and an escape time evaluation, which ties the control back into the choice of quantizer saturation bound. An example is studied which illustrates the role of the control objective in determining the efficacy of coding using these schemes. The results mesh well with those observed in signal coding. However, the introduction of a realization-based escape time is a novelty departing significantly from mean square computations.

4.1 Introduction

We consider a linear plant with input u_t and output y_t connected to a controller by a noise-free fixed-bitrate-b memoryless channel. The measured output is coded for transmission through the channel and we consider several period-two coding or bitrate assignment strategies. In each case, the output is quantized with a linear fixed quantizer with saturation bound ζ. The coding strategies perform a period-two bit-allocation for the signal being communicated across the channel. In Strategy I, the b bits of a b-bit quantizer are sent at each instant. Strategy II applies a $2b$-bit quantizer and sends alternately the most significant b bits and the least significant b bits of the even-timed output sample. The strategies differ in their delays and accuracy; y_t has b bits at each time versus y_{2t} has b bits at time $2t$ and $2b$ bits at time $2t+1$. No information is transmitted about y_{2t+1} in the second strategy. A third, intermediate strategy is also examined. These coding/bit-assignment schemes are evaluated using the LQ performance of the controlled plant. Using a result from Curry [1], the optimal control will comprise linear state-variable feedback and a conditional mean estimator using the decoded output.

A quantizer is the functional composition or cascade of two distinct memoryless characteristic; an infinite quantizer and a saturation. This is depicted in Figure 4.1. We divide our

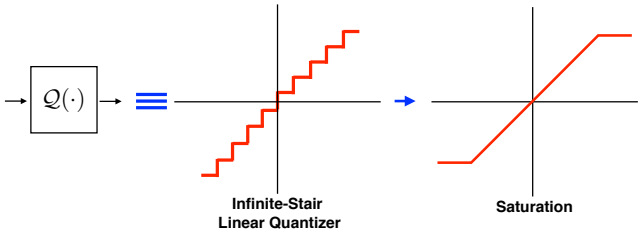

Figure 4.1. Representation of a quantizer as the functional composition of two memoryless nonlinearities; an infinite quantizer and a saturation. The analysis treats each component in succession.

analysis to consider each nonlinear aspect separately. In the case where the quantizer does not saturate and we use subtractive dithered quantization, the optimal conditional mean estimator is the Kalman filter, whose state estimate error covariance is computable using standard methods.

The quantizer step size appears in the measurement noise term. For each coding strategy this covariance is simply computed and the LQ control performance derived. Next , the saturation nonlinearity is introduced by using these second-order signal statistics to compute the expected time before saturation. This *escape time* is a function of the closed-loop controlled signal y_t. We use this property to tie to the controller objective function to the selection of the quantizer bound, ζ. For a given feasible set of escape time, bitrate and control objective, there is a saturation bound and LQ performance. As the coding strategies change, so too does this performance. For a given escape time, we compare the quasi-stationary performance.

Contribution of this paper

- By treating a limited set of coding schemes, we are able to draw conclusions about coding in the output signal path. The range and the correlation/predictability of the closed-loop plant output play a role in the efficacy of coding. Less predictable outputs, such as those of minimum variance control, benefit less from coding. This concurs with observations in signal processing.
- At low bitrates coding can become important.
- The decomposition of the quantizer into two factors admits analysis using the escape time ideas introduced in this paper. This makes the study of methods possible by separating the estimator performance from the saturation behavior.
- The escape time analysis permits the consideration of stabilization problems and performance together. The focus on realization based behavioral descriptors admits new viewpoints compared with asymptotic moments.

Relevant prior work

Borkar and Mitter [42] study a full-state feedback formulation with vector quantization and coding delay similar to the strategies in this paper. They use the full state measurement

to compute the process noise and then code this using vector quantization. They define a delay-accuracy tradeoff denoted by N and indicating the number of noise samples held before transmission. Longer delay admits multiples of the underlying bitrate when eventually transmitted. When $N = 1$, their results are similar to our Strategy I and when $N = 2$ they resemble our Strategy II. An optimal vector quantizer [43] then encodes the data into the available bits. This vector quantizer yields the conditional mean process noise reconstruction at the receiver even though the channel is not error-free. The decoded value is then used to construct the conditional mean state estimate. Then could apply the same theorem from Curry in [1] to which we appeal shortly. By limiting the discussion to stable systems, as in our earlier paper [44], they are able to avoid explicit discussion of the saturation issues with quantization, vector or otherwise.

Fu in [45] studied the coding problem for the control signal of linear quadratic Gaussian control with a memoryless error-free channel of fixed rate. The paper deals with optimization over the set of causal encoders and their decoder pairs. Fu looks only at finite-horizon optimal control and therefore sidesteps the stability and saturation questions. He does, however, develop a value for the finite-horizon LQG performance, which involves the distortion function connects the coder and the objective function. In practice, optimizing this distortion appears intractable. He presents in Theorem 4.1, a corrected version of Fischer's result [46], a *weak* separation theorem where the estimate distortion function D is a function of the control problem and not just the estimation problem parameters. When he considers the optimal coding problem for even a simple initial condition case, the solution depends on the control objective and the effect of the current encoding on future distortions. So his coder needs both memory and look-ahead and the problem begins to mimic the intractability of stochastic optimal control. In the current paper, by limiting our discussion to specific coding strategies, we reveal other aspects of a complicated picture. By limiting our coders to being memoryless, we are able to appeal to the separation theorem of Curry.

Nair and Evans [47] treat adaptive coding to achieve stabilization with limited bitrate. They assign on level of an adaptive quantizer to indicate saturation. When this level is received at

the decoder, the quantizer range is expanded multiplicatively. Effectively, the bitrate required to stabilize an unstable system is tied to being able to achieve the expansion as a sufficiently rapid rate to catch the unstable output. This imaginative coding scheme concentrates on stabilization in mean square and does not address signal limits nor controlled performance.

The impact of quantization on performance at high rates is explored in [48], the state of the system being quantized prior to transmission to the controller and they assess the performance of the controller to minimize a quadratic cost.

A similar approach is explored in [49] pertinent for speech coding but close to the current paper, in particular Strategies I and II. These strategies are applied to speech with an autoregressive model. The performance is evaluated qualitatively by Mean Opinion Score. They show that down-sampling plus smoothing leads to better coding results for highly-correlated voiced speech and low-delay coding is preferable for unvoiced speech, which resembles modulated white noise.

Kostina and Hassibi [50] consider LQR optimal control and the question of minimal channel capacity required to achieve a given bound, b, on the expected LQR cost. They explore the problem with fully observed and partially observed state. In addition to this capacity bound, they explore specific lattice codes which achieve the bound. The bound itself depends on both control and estimation aspects for the partially observed case. They consider an error-free channel and explore all possible causal codes. Their communications structure is a limited capacity forward channel from the transmitter to receiver/controller together with a side channel which conveys the controller's state prediction back to the encoder. The minimizing codes transmit quantized versions of the error between the transmitter's state (or state prediction) and the receiver's state prediction. This communications structure obviates the requirement for the system to be stable. Although, similarly to [47], the logarithm of the determinant of the system matrix appears in the capacity bound.

4.2 Problem statement

Consider the following optimal control problem.

- Linear plant system with Gaussian noises:

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad x_0, \tag{4.1}$$

$$y_t = Cx_t + v_t, \tag{4.2}$$

  Here, state $x_t \in \mathbb{R}^n$, input $u_t \in \mathbb{R}^p$, output $y_t \in \mathbb{R}^m$, process noise $w_t \in \mathbb{R}^n$, measurement noise $v_t \in \mathbb{R}^m$. Noise sequences $\{w_t\}$ and $\{v_t\}$ are Gaussian, mutually independent, zero-mean and white with known covariances. The plant initial condition is also Gaussian and independent from w_t and v_t for all t.

- Quadratic performance criterion, minimized over non-anticipatory controls, u_t, computed from the received data at the controller,

$$J_{LQ} = \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left(\sum_{t=1}^{N} x_t^T Q_c x_t + u_t^T R_c u_t \right).$$

- The communications link between the plant measurement and the control computation consists of a limited bitrate, b-bits-per-sample, memoryless noise-free channel.

- The coder-controller is restricted to the following elements.

- The measurement y_t is quantized to a fixed number of bits, which can be larger than b.

- Some of these bits are encoded into the bitstream forwarded to the controller subject to the bitrate limit.

- For this paper, we restrict attention to period-two bitrate assignment strategies.

- The controller computes and applies the control.

4.2.1 LQ Optimal Controller

Denote by $\{p_t\}$ the sequence of decoded signal values available at the controller. Then, we have the following result from Curry.

Theorem 17 (Curry [1]). *For the linear state system (4.1)-(4.2) with nonlinear, memoryless measurement*

$$p_t = \varphi_t(x_t, v_t),$$

with $\{v_t\}$ *white and independent from* x_t *and quadratic objective function*

$$J_t = E\left(\sum_{k=t}^{N+1} x_k^T Q_k x_k + u_k^T R_k u_k \middle| \mathbf{P}^t, \mathbf{U}^{t-1}, \pi_{0|-1} \right), \tag{4.3}$$

the optimal output feedback control is given by

$$u_t^\star = -K_t E(x_t | \mathbf{P}^t, \mathbf{U}^{t-1}, \pi_{0|-1}),$$

where, $\mathbf{P}^t = \{p_1, p_2, ..., p_t\}$, $\mathbf{U}^t = \{u_1, u_2, ..., u_t\}$ *and* K_t *is the standard LQ optimal feedback gain.*

Decoding signal p_t at the receiver side, the filtered plant state estimate $\hat{x}_{t|t}$ and infinite-horizon control law $u_t = -K\hat{x}_{t|t}$ are computed with $K = \text{dare}(A, B, Q_c, R_c)$. The performance is evaluated with the LQ criterion. Signal p_t will be derived from output y_t by quantization and coding.

4.3 Controller Coding Strategies

The problem statement imposes the quantization of plant output signal y_t. We restrict our attention to uniform quantization and limit consideration to subtractive dithered quantizers in order to facilitate the receiver-side estimation.

4.3.1 Dithered Quantization

A *subtractive b-bit dithered quantizer,* $\mathbf{Q}_b(\cdot)$, is a memoryless function which takes input signal y_t and dither signal, d_t, and produces an output signal

$$\mathbf{Q}_b(y_t) = Q_b(y_t + d_t) - d_t, \tag{4.4}$$

where $Q_b(\cdot)$ is a standard uniform quantizer. Such quantizers are examined in detail in, for example, [3].

Theorem 18. *Consider a uniform, midrise, symmetric, b-bits-per-channel, subtractive dithered quantizer,* $\mathbf{Q}_b(\cdot)$*, with saturation bounds* $\pm\zeta$*. Assume:*

(A) dither d_t *is a white noise process independent from* y_t *with a probability density possessing characteristic function,* $\Phi_d(\cdot)$*, satisfying* $\Phi_d\left(l\frac{\pi 2^b}{\zeta}\right) = 0$ *for* $l = \pm 1, \pm 2, \ldots,$

(B) $y_t + d_t \in [-\zeta, \zeta]$*, i.e. no saturation of the dithered quantizer occurs.*

Then, the quantization error

$$q_{b,t} \triangleq \mathbf{Q}_b(y_t) - y_t, \tag{4.5}$$

is: (i) white, (ii) independent from y_t*, (iii) uniformly distributed on* $\left[-\frac{\zeta}{2^b}, \frac{\zeta}{2^b}\right]$.

This theorem, an embellishment of Theorem QTSD of [3], presents conditions under which the quantization error is an additive white noise independent from the signal being quantized as studied with details in [2]. Denote the quantizer step size as

$$\Delta = \frac{\zeta}{2^{b-1}}.$$

Then, we note the following.

$$q_{b,t} \sim \mathscr{U}\left[-\frac{\Delta}{2},\frac{\Delta}{2}\right],\ \mathrm{E}(q_{b,t})=0,$$
$$\mathrm{cov}(q_{b,t}) = \frac{\zeta^2}{3\times 2^{2b}} \triangleq S_b. \tag{4.6}$$

We also, note that the characteristic function condition is satisfied by dither which is uniformly distributed $\mathscr{U}[-\Delta/2,\Delta/2]$ or which is triangularly distributed tr$[-\Delta,\Delta]$, for example. In our calculations later, we use uniform dither d_t.

4.3.2 Period-two Bit-assignment and Transmission Strategies

We consider a fixed-rate, b-bits-per-transmission, channel and propose three period-two quantization strategies which reflect similar approaches from Signal Processing [51], [52], [49]. The intention is to manage the quantization error with periodic changes to the effective bitrate and allied signal delay. We will examine the efficacy of these methods in terms of their benefits for LQ output feedback control.

The presence of the b-bits-per-sample channel militates that the subtractive dithered quantizer operates on both sides of the channel. That is, b bits are transmitted each sample as symbol m_k from the transmitter. Then at the receiver subtractive dither is applied. This and other implementation issues of wordlength etc. are discussed in [3]. With our period-two strategies, both the dithering and the subtraction will be modified. Here $\mathrm{MSB}_n(x_t)$ and $\mathrm{LSB}_n(x_t)$ denote the most significant and least significant n bits of signal x_t. While m_t is the b-bit transmitted message at time t, p_t or p'_t denotes the reconstructed/decoded plant output at the receiver for input into the Kalman filter.

Strategy I

1: **for** t even or odd **do**

2: $m_t = Q_b(y_t + d_t^b)$ is transmitted

3: $p_t \leftarrow m_t - d_t^b$ at the receiver

$\hat{x}_{t|t} \leftarrow$ (4.7) Kalman filter Lemma 19

$u_t \leftarrow -K\hat{x}_{t|t}$

Strategy II

if $t = 2k$, even time, **then**

$m_{2k} = \mathrm{MSB}_b(Q_{2b}(y_{2k} + d_{2k}^{2b}))$ is transmitted

$p_{2k} \leftarrow m_{2k}$ at the receiver without dither subtraction

$\hat{x}_{2k|2k} \leftarrow$ (4.8) Kalman filter from Lemma 20

$u_{2k} \leftarrow -K\hat{x}_{2k|2k}$

else $t = 2k+1$, odd time,

$m_{2k+1} = \mathrm{LSB}_b(Q_{2b}(y_{2k} + d_{2k}^{2b}))$ is transmitted

$p_{2k+1} \leftarrow p_{2k} + 2^{-b} m_{2k+1} - d_{2k}^{2b}$ at the receiver

$\hat{x}_{2k+1|2k+1} \leftarrow$ (4.9) Kalman filter from Lemma 20

$u_{2k+1} \leftarrow -K\hat{x}_{2k+1|2k+1}$

Strategy III

if $t = 2k$, even time, **then**

$m_{2k} = \mathrm{MSB}_b(Q_{b+r}(y_{2k} + d_{2k}^{b+r}))$ is transmitted

$p_{2k} \leftarrow m_{2k}$ at the receiver without dither subtraction

$\hat{x}_{2k|2k} \leftarrow$ (4.10) Kalman filter from Lemma 21

$u_{2k} \leftarrow -K\hat{x}_{2k|2k}$

else $t = 2k+1$, odd time,

$m_{2k+1} = \mathrm{LSB}_r(Q_{b+r}(y_{2k} + d_{2k}^{b+r}))$

$\qquad + 2^{-r}\mathrm{MSB}_{b-r}(Q_{b-r}(y_{2k+1} + d_{2k+1}^{b-r}))$

is transmitted

$p'_{2k} \leftarrow p_{2k} + 2^{-b}\mathrm{MSB}_r(m_{2k+1}) - d_{2k}^{b+r}$

$p_{2k+1} \leftarrow \mathrm{LSB}_{b-r}(m_{2k+1}) - d_{2k+1}^{b-r}$

$\hat{x}_{2k+1|2k+1} \leftarrow$ (4.11) Kalman filter from Lemma 21

13: $u_{2k+1} \leftarrow -K\hat{x}_{2k+1|2k+1}$

We note two central features of the time-varying strategies.

- Strategy I uses a quantizer and associated dither of step size $\frac{\zeta}{2^{b-1}}$ while Strategy II uses step size $\frac{\zeta}{2^{2b-1}}$, and Strategy III uses alternately $\frac{\zeta}{2^{2(b+r)-1}}$ and $\frac{\zeta}{2^{2r-1}}$.

- Strategies II and III at even times receive undithered b-most-significant-bit transmissions, since the dither operates further along the bitstream. Accordingly, the quantization error at even times is not white, nor uniform, nor independent from y_{2k}, even though the quantization noise for y_{2k} at time $2k+1$ does possess these properties. We shall conduct our analysis blithely without taking these even quantization error properties fully into account.

We note that, with Strategies II and III, the state estimate calculation will be non-standard at the controller, reflecting the periodic information pattern. The associated Kalman filter will be presented shortly and computes $\hat{x}_{2k|2k+1}$ and then $\hat{x}_{2k+1|2k+1}$ from the received data. The derivation of these filters and their properties is a core contribution of the paper.

4.4 Kalman Filters and Covariances for the Strategies

We derive the Kalman filters associated with each of the strategies under the following assumption.

Assumption 3 (For this and the following sections alone)**.** *The quantizer never saturates. That is, $y_t + d_t \in [-\zeta, \zeta]$. So, following Theorem 18, the quantization errors:*

Strategy I: $p_t - y_t$*;*

Strategy II: $p_{2k+1} - y_{2k}$*;*

Strategy III: $p'_{2k} - y_{2k}$ *and* $p_{2k+1} - y_{2k+1}$*;*

are independent from $\{y_t\}$*, white, zero-mean, uniformly distributed with covariances* S_b*,* S_{2b}*,* S_{b+r} *and* S_{b-r} *respectively, where* S_b *is defined in* (4.6)*.*

Further and without justification, we assume that the other quantization errors, $p_{2k} - y_{2k}$ *in Strategies II and III, satisfy the same properties with covariances* S_b*.*

Assumption 4. *Each strategy's state estimator commences with state estimate* $\hat{x}_{0|-1}$ *and covariance* $\Sigma_{0|-1}$*, at* $t = 0$*.*

The following results for Strategies I, II and III are proved in the Appendix.

Lemma 19 (Anderson, Moore [29])**.** *For Strategy I, the Kalman filter driven by signal* p_t *from Algorithm I Line 3 is calculated by:*

$$\begin{aligned} L_t &= \Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + R + S_b)^{-1}, \\ \hat{x}_{t|t} &= \hat{x}_{t|t-1} + L_t(p_t - C\hat{x}_{t|t-1}), \\ \hat{x}_{t+1|t} &= (A - BK)\hat{x}_{t|t}, \\ \Sigma_{t+1|t} &= A\Sigma_{t|t-1}A^T - AL_tC\Sigma_{t|t-1}A^T + Q. \end{aligned} \tag{4.7}$$

Lemma 20. *For Strategy II, the Kalman filter driven by signals* p_{2k} *from Algorithm II Line 3 and* p_{2k+1} *from Line 8 is calculated by:*

At even times, $t = 2k$ *:*

$$\begin{aligned} L_{2k} &= \Sigma_{2k|2k-1}C^T(C\Sigma_{2k|2k-1}C^T + R + S_b)^{-1}, \\ \hat{x}_{2k|2k} &= \hat{x}_{2k|2k-1} + L_{2k}(p_{2k} - C\hat{x}_{2k|2k-1}), \end{aligned} \tag{4.8}$$

At odd times, $t = 2k+1$:

$$
\begin{aligned}
L_{2k+1} &= \Sigma_{2k|2k-1} C^T (C\Sigma_{2k|2k-1} C^T + R + S_{2b})^{-1}, \\
\hat{x}_{2k+1|2k+1} &= A\left(\hat{x}_{2k|2k-1} + L_{2k+1}(p_{2k+1} - C\hat{x}_{2k|2k-1})\right) \\
&\qquad - BK\hat{x}_{2k|2k}, \\
\hat{x}_{2k+2|2k+1} &= (A - BK)\hat{x}_{2k+1|2k+1}, \\
\Sigma_{2k+2|2k+1} &= A^2 \Sigma_{2k|2k-1} A^{2^T} - A^2 L_{2k+1} C \Sigma_{2k|2k-1} A^{2^T} \\
&\qquad + AQA^T + Q.
\end{aligned}
\tag{4.9}
$$

Lemma 21. *For Strategy III, the Kalman filter driven by signals p_{2k} from Algorithm III Line 3, p'_{2k} Line 10 and p_{2k+1} Line 11 is calculated by:*

At even times, $t = 2k$,

$$
\begin{aligned}
L_{2k} &= \Sigma_{2k|2k-1} C^T (C\Sigma_{2k|2k-1} C^T + R + S_b)^{-1}, \\
\hat{x}_{2k|2k} &= \hat{x}_{2k|2k-1} + L_{2k}(p_{2k} - C\hat{x}_{2k|2k-1}),
\end{aligned}
\tag{4.10}
$$

At odd times, $t = 2k+1$,

$$
\begin{aligned}
L'_{2k} &= \Sigma_{2k|2k-1} C^T (C\Sigma_{2k|2k-1} C^T + R + S_{b+r})^{-1}, \\
\hat{x}'_{2k+1|2k} &= A\hat{x}_{2k|2k-1} + AL'_{2k}(p'_{2k} - C\hat{x}_{2k|2k-1}) \\
&\qquad\qquad - BK\hat{x}_{2k|2k}, \\
\Sigma'_{2k+1|2k} &= A\Sigma_{2k|2k-1}A^T - A\Sigma_{2k|2k-1}C^T \times \\
&\quad (C\Sigma_{2k|2k-1}C^T + R + S_{b+r})^{-1} C\Sigma_{2k|2k-1}A^T + Q, \\
L_{2k+1} &= \Sigma'_{2k+1|2k} C^T (C\Sigma'_{2k+1|2k} C^T + R + S_{b-r})^{-1}, \\
\hat{x}_{2k+1|2k+1} &= \hat{x}'_{2k+1|2k} + L_{2k+1}(p_{2k+1} - C\hat{x}'_{2k+1|2k}), \\
\hat{x}_{2k+2|2k+1} &= (A - BK)\hat{x}_{2k+1|2k+1}, \\
\Sigma_{2k+2|2k+1} &= A\Sigma'_{2k+1|2k}A^T - A\Sigma'_{2k+1|2k}C^T \times \\
&\quad (C\Sigma'_{2k+1|2k}CT + R + S_{b-r})^{-1} \Sigma'_{2k+1|2k}A^T + Q.
\end{aligned}
\tag{4.11}
$$

Although both covariances $\lim_{k\to\infty}\Sigma_{2k|2k-1}$ and $\lim_{k\to\infty}\Sigma_{2k+1|2k}$ may have different limiting values for Strategies II & III, the value of the former suffices for the rest of the calculation.

Corollary 22. *For Strategy I, $\Sigma_I^{p,\infty} \triangleq \lim_{k\to\infty}\Sigma_{k|k-1}$ and $\Sigma_I^{\infty} \triangleq \lim_{k\to\infty}\Sigma_{k|k}$ satisfy*

$$
\begin{aligned}
\Sigma_I^{p,\infty} &= \texttt{dare}(A^T, C^T, Q, R + S_b), \\
\Sigma_I^{\infty} &= \Sigma_I^{p,\infty} - \Sigma_I^{p,\infty} C^T (C\Sigma_I^{p,\infty} C^T + R + S_b)^{-1} C\Sigma_I^{p,\infty}.
\end{aligned}
\tag{4.12}
$$

Corollary 23. *For Strategy II, $\Sigma_{II}^{p,\infty} \triangleq \lim_{k\to\infty}\Sigma_{2k|2k-1}$, $\Sigma_{II_{even}}^{\infty} \triangleq \lim_{k\to\infty}\Sigma_{2k|2k}$ and $\Sigma_{II_{odd}}^{\infty} \triangleq$*

$\lim_{k\to\infty}\Sigma_{2k+1|2k+1}$ *satisfy*

$$\begin{aligned}
\Sigma_{II}^{p,\infty} &= \texttt{dare}\left(A^{2^T}, C^T, AQA^T+Q, R+S_{2b}\right),\\
\Sigma_{II_{even}}^{\infty} &= \Sigma_{II}^{p,\infty} - \Sigma_{II}^{p,\infty}C^T(C\Sigma_{II}^{p,\infty}C^T+R+S_b)^{-1}C\Sigma_{II}^{p,\infty},\\
\Sigma_{II_{odd}}^{\infty} &= \Sigma_{II}^{p,\infty} - \Sigma_{II}^{p,\infty}C^T(C\Sigma_{II}^{p,\infty}C^T+R+S_{2b})^{-1}C\Sigma_{II}^{p,\infty}.
\end{aligned} \tag{4.13}$$

Corollary 24. *For Strategy III,* $\Sigma_{III}^{p,\infty}\lim_{k\to\infty} \triangleq \Sigma_{2k|2k-1}$, $\Sigma_{III_{even}}^{\infty} \triangleq \lim_{k\to\infty}\Sigma_{2k|2k}$, *and* $\Sigma_{III_{odd}}^{\infty} \triangleq \lim_{k\to\infty}\Sigma_{2k+1|2k+1}$, *satisfy*

$$\begin{aligned}
\Sigma_{III}^{p,\infty} &= \texttt{dare}\left(A^{2^T}, \mathscr{G}_1, AQA^T+Q, \mathscr{G}_2, \begin{bmatrix}0 & AQC^T\end{bmatrix}, \texttt{eye}(n)\right),\\
\Sigma_{III_{even}}^{\infty} &= \Sigma_{III}^{p,\infty} - \Sigma_{III}^{p,\infty}C^T(C\Sigma_{III}^{p,\infty}C^T+R+S_{b+r})^{-1}C\Sigma_{III}^{p,\infty},\\
\Sigma_{III_{odd}}^{\infty} &= \Sigma_{III}^{p,\infty} - \Sigma_{III}^{p,\infty}C^T(C\Sigma_{III}^{p,\infty}C^T+R+S_{b-r})^{-1}C\Sigma_{III}^{p,\infty},
\end{aligned} \tag{4.14}$$

where

$$\mathscr{G}_1 = \begin{bmatrix}C^T & A^TC^T\end{bmatrix},\ \mathscr{G}_2 = \begin{bmatrix}R+S_{b+r} & 0\\ 0 & CQC^T+R+S_{b-r}\end{bmatrix}.$$

4.5 Control performance analysis

The limiting performance of three strategies may be computed using standard covariance methods.

Definition 1. *The* i,j*-block* $(n\times n)$ *entry of matrices* Ψ_X*, below is denoted by* $\Psi_X(i,j)$ *for* $X = I, II, III$.

Theorem 25. *Subject to Assumption 3, the performance for Strategy I given by*

$$J_I = trace[Q_c\Psi_I(1,1)] + trace[K^TR_cK\Psi_I(2,2)], \tag{4.15}$$

calculated through these steps:

(i) $K = \texttt{dare}(A, B, Q_c, R_c)$,

(ii) $\Sigma_I^{p,\infty} = \texttt{dare}\left(A^T, C^T, Q, R + S_b\right)$,

(iii) $L = \Sigma_I^{p,\infty} C^T \left(C\Sigma_I^{p,\infty} C^T + R + S_b\right)^{-1}$,

(iv) $\Psi_I = \texttt{dlyap}\left(\mathscr{M}_1, \mathscr{N}_1 \mathscr{P}_1 \mathscr{N}_1^T\right)$,

where

$$\mathscr{M}_1 = \begin{bmatrix} A & -BK \\ LCA & (I-LC)A - BK \end{bmatrix}, \quad \mathscr{N}_1 = \begin{bmatrix} I & 0 & 0 \\ LC & L & L \end{bmatrix},$$

$$\mathscr{P}_1 = \begin{bmatrix} Q & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & S_b \end{bmatrix}, \quad \Psi_I = \begin{bmatrix} E(x_k x_k^T) & E(x_k \hat{x}_{k|k}^T) \\ E(\hat{x}_{k|k} x_k^T) & E(\hat{x}_{k|k} \hat{x}_{k|k}^T) \end{bmatrix}.$$

Theorem 26. *Subject to Assumption 3, the performance for Strategy II, given by*

$$\begin{aligned} J_{II} = &\frac{1}{2} trace\left\{Q_c[\Psi_{II}(1,1) + \Psi_{II}(3,3)]\right\} + \\ &\frac{1}{2} trace\left\{K^T R_c K[\Psi_{II}(2,2) + \Psi_{II}(4,4)]\right\}, \end{aligned} \tag{4.16}$$

calculated through these steps:

(i) $K = \texttt{dare}(A, B, Q_c, R_c)$,

(ii) $\Sigma_{II}^{p,\infty} = \texttt{dare}\left(A^{2^T}, C^T, AQA^T + Q, R + S_{2b}\right)$,

(iii) $L_{even} = \Sigma_{II}^{p,\infty} C^T \left(C\Sigma_{II}^{p,\infty} C^T + R + S_b\right)^{-1}$.

(iv) $L_{odd} = \Sigma_{II}^{p,\infty} C^T (C\Sigma_{II}^{p,\infty} C^T + R + S_{2b})^{-1}$,

(v) $\Psi_{II} = \texttt{dlyap}\left(\mathscr{M}_2, \mathscr{N}_2 \mathscr{P}_2 \mathscr{N}_2^T\right)$

where

$$\mathscr{M}_2 = F_4F_3F_2F_1,\ \ \mathscr{N}_2 = \begin{bmatrix} F_4F_3F_2G_1 & F_4G_3 & F_4F_3G_2 & G_4 \end{bmatrix},$$

$$\mathscr{P}_2 = \begin{bmatrix} Q & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & R+S_b & R+S_{2b} \\ 0 & 0 & R+S_{2b} & R+S_{2b} \end{bmatrix},$$

$$\Psi_{II} = E\left(\begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k} \\ x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \end{bmatrix} \begin{bmatrix} x_{2k}^T & \hat{x}_{2k|2k}^T & x_{2k+1}^T & \hat{x}_{2k+1|2k+1}^T \end{bmatrix} \right)$$

$$G_1 = \begin{bmatrix} 0 \\ I \end{bmatrix},\ \ G_2 = \begin{bmatrix} 0 \\ 0 \\ L_0 \end{bmatrix},\ \ G_3 = \begin{bmatrix} 0 \\ 0 \\ L_1 \end{bmatrix},\ \ G_4 = \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \end{bmatrix}$$

$$F_1 = \begin{bmatrix} 0 & 0 & 0 & A-BK \\ 0 & 0 & A & -BK \end{bmatrix},\ \ F_2 = \begin{bmatrix} I & 0 \\ 0 & I \\ (I-L_0C) & L_0C \end{bmatrix},$$

$$F_3 = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ (I-L_1C) & L_1C & 0 \end{bmatrix},\ \ F_4 = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ A & -BK & 0 \\ 0 & -BK & A \end{bmatrix}.$$

Note, the two-step update is described by the recursion

$$\begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k} \\ x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \end{bmatrix} = \mathscr{M}_2 \begin{bmatrix} x_{2k-2} \\ \hat{x}_{2k-2|2k-2} \\ x_{2k-1} \\ \hat{x}_{2k-1|2k-1} \end{bmatrix} + \mathscr{N}_2 \begin{bmatrix} w_{2k-1} \\ w_{2k} \\ \begin{bmatrix} v_{2k}+q_{2k} \\ v_{2k}+q_{2k+1} \end{bmatrix} \end{bmatrix}.$$

Theorem 27. *Subject to Assumption 3, the performance for Strategy III, given by*

$$\begin{aligned} J_{III} = &\frac{1}{2} trace\left\{Q_c[\Psi_{III}(1,1)+\Psi_{III}(3,3)]\right\}+ \\ &\frac{1}{2} trace\left\{K^T R_c K[\Psi_{III}(2,2)+\Psi_{III}(4,4)]\right\}, \end{aligned} \tag{4.17}$$

calculated through these steps:

(i) $K = \texttt{dare}(A,B,Q_c,R_c)$,

(ii) $\Sigma^{p,\infty} = \texttt{dare}\left(A^{2^T}, \mathscr{G}_1, AQA^T+Q, \mathscr{G}_2, \begin{bmatrix} 0 & AQC^T \end{bmatrix}, \texttt{eye}(n)\right)$,

(iii) $L_{even} = \Sigma^{p,\infty}C^T\left(C\Sigma^{p,\infty}C^T+R+S_b\right)^{-1}$,

(iv) $L_{odd1} = \Sigma^{p,\infty}C^T(C\Sigma^{p,\infty}C^T+R+S_{b+r})^{-1}$,

(v) $L_{odd2} = \Sigma^{p,\infty}C^T(C\Sigma^{p,\infty}C^T+R+S_{b-r})^{-1}$,

(vi) $\Psi_{III} = \texttt{dlyap}\left(\mathscr{M}_3, \mathscr{N}_3\mathscr{P}_3\mathscr{N}_3^T\right)$,

where

$$\mathscr{M}_2 = F_4F_3F_2F_1,$$

$$\mathscr{N}_2 = \begin{bmatrix} F_4F_3F_2G_1 & F_4G_3 & F_4F_3G_2 & G_4 \end{bmatrix},$$

$$\mathscr{P}_3 = \begin{bmatrix} Q & 0 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 & 0 \\ 0 & 0 & R+S_b & R+S_{b+r} & 0 \\ 0 & 0 & R+S_{b+r} & R+S_{b+r} & 0 \\ 0 & 0 & 0 & 0 & R+S_{b-r} \end{bmatrix}$$

$$\Psi_{III} = E\left(\begin{bmatrix} x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \\ x_{2k} \\ \hat{x}^1_{2k|2k} \end{bmatrix} \begin{bmatrix} x^T_{2k+1} & \hat{x}^T_{2k+1|2k+1} & x^T_{2k} & \hat{x}^{1^T}_{2k|2k} \end{bmatrix} \right)$$

$$G_1 = \begin{bmatrix} I \\ 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & 0 \\ L_{even} & 0 \\ 0 & L_{odd1} \end{bmatrix}, \quad G_3 = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad G_4 = \begin{bmatrix} 0 \\ L_{odd2} \\ 0 \\ 0 \end{bmatrix},$$

$$F_1 = \begin{bmatrix} A & -BK & 0 & 0 \\ 0 & A-BK & 0 & 0 \end{bmatrix}, \quad F_2 = \begin{bmatrix} I & 0 \\ L_{even}C & (I-L_{even}C) \\ L_{odd1}C & (I-L_{odd1}C) \end{bmatrix}, \quad F_3 = \begin{bmatrix} A & -BK & 0 \\ 0 & 0 & I \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix},$$

$$F_4 = \begin{bmatrix} I & 0 & 0 & 0 \\ L_{odd2}C & (I-L_{odd2}C)A & 0 & -(I-L_{odd2}C)BK \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix},$$

The two-step update is described by the recursion

$$\begin{bmatrix} x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \\ x_{2k} \\ \hat{x}^1_{2k|2k} \end{bmatrix} = \mathscr{M}_3 \begin{bmatrix} x_{2k-1} \\ \hat{x}_{2k-1|2k-1} \\ x_{2k-1} \\ \hat{x}^1_{2k-2|2k-2} \end{bmatrix} + \mathscr{N}_3 \begin{bmatrix} w_{2k-1} \\ w_{2k} \\ \begin{bmatrix} v_{2k} + q_{b,2k} \\ v_{2k} + q_{b+r,2k} \end{bmatrix} \\ v_{2k+1} + q_{b-r,2k+1} \end{bmatrix}.$$

4.6 Escape time analysis

The performance analysis from earlier sections is based on direct second moment calculations subject to the validity of Assumption 3, i.e. that the controlled system output

$$z_t = y_t + d_t,$$

satisfies $|z_t| \leq \zeta$. For Gaussian y_t, or indeed for any y_t with density of unbounded support, the signal $y_t + d_t$ is guaranteed to exceed this bound infinitely often. Our aim in this section is to quantify the average residence time of the dithered controlled output signal inside the saturation bound. If this residence time is long, then the earlier linear analysis will remain valid on average for a long time and can be used to characterize performance, since the stabilizing control yields a quasi-stationary closed loop subject to no saturation. This will be validated by computational experiments in Section 4.7.

We make the following definition.

Definition 2. *The escape time, τ_{esc}, is the first time that $z_t \notin [-\zeta, \zeta]$.*

Our aim is now to calculate the mean escape time as a function of ζ. This will demonstrate that the choice of ζ to yield a particular mean escape time depends on the choice of state feedback control gain K. The state estimation covariance analysis of Section 4.4 did not explicitly depend on K. But now, via its effect on ζ, the control problem affects this covariance.

If we have ergodicity of the stochastic process $\{z_t\}$ then the long-term sample average frequency of z_t falling outside $[-\zeta, \zeta]$ is equal to the ensemble average computable from the density of z_t. If the Gaussian process $\{y_t\}$ is ergodic, then since $\{d_t\}$ is white and stationary, the signal $\{y_t + d_t\}$ is ergodic. We have the following theorem from Caines [53] who cites earlier sources going back to Maruyama and Grenander.

Theorem 28. *[53] A necessary and sufficient condition for a discrete-time stationary Gaussian process to be ergodic is that the spectral distribution of the process is continuous.*

If y_t is the output of a stable linear system driven by white, independent, zero-mean Gaussian noises n_t and r_t with covariances Q and R respectively, that is,

$$\xi_{t+1} = F\xi_t + Gn_t,$$
$$y_t = Hp_t + Jr_t,$$

then its power spectral density is given by

$$\Phi_{yy}(\omega) = JRJ^T + H(e^{j\omega}I - F)^{-1}GQG^T(e^{-j\omega}I - F^T)^{-1}.$$

If the eigenvalues of F are within $|z| < 1$ and $JRJ^T > 0$, then y_t is ergodic by Theorem 28, as is the signal z_t.

For our LQG problem, y_t is generated with

$$F = \begin{bmatrix} A & -BK \\ LCA & A-BK-LCA \end{bmatrix}, \quad G = \begin{bmatrix} I & 0 \\ LC & L \end{bmatrix},$$
$$H = \begin{bmatrix} C & 0 \end{bmatrix}, \quad J = I,$$

which F has all eigenvalues inside the unit circle by construction subject to the conditions in the following theorem.

Theorem 29. *[54] Subject to Assumption 3, provided $R_c > 0$, $R > 0$, $[A, Q_c]$ detectable, $[A, Q]$ stabilizable, the dithered controlled output signal, $z_t = y_t + d_t$, is asymptotically stationary and ergodic. So*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{|z_t|>\zeta} = Pr(|z_t| > \zeta), \tag{4.18}$$

where $\mathbb{1}_A$ is the indicator function of event A.

Once we have ergodicity of the closed-loop signal z_t, then we have the following result.

Theorem 30. *For ergodic z_t, if $Pr(|z_k| > \zeta) = \beta$, then the expected escape time is $E[\tau_{esc}] = \frac{1}{\beta}$.*

These are the steps and important parameters of the analysis.

1. Choose a desired average escape time $\mathrm{E}[\tau_{\mathrm{esc}}]$. The escape probability is $\beta = \frac{1}{\mathrm{E}[\tau_{\mathrm{esc}}]}$.
2. Initiate the line search for ζ depending on coding strategy and using one of
   - $\Psi_I(1,1)$ from (4.15), or
   - $\Psi_{II}(1,1)$ and $\Psi_{II}(3,3)$ from (4.16), or
   - $\Psi_{III}(1,1)$ and $\Psi_{III}(3,3)$ from (4.17),

   compute Z, the covariance of z_t. Then solve

$$\frac{\beta}{2} = \texttt{mvncdf}\left(-\zeta_{\mathrm{new}}.\mathbb{1}_m, 0_m, Z\right),$$

where `mvncdf` is the multivariate normal cumulative distribution function.

4.7 Numerical examples

We compare coding strategies in the following examples through certain steps.

1. With given $\{A, B, C, Q, R, Q_c, R_c\}$, compute the linear feedback gain K, via Theorem 25 Step (i).

2. Fix the mean escape time, τ_{esc}.

3. For each coding strategy, compute the corresponding quantizer bound, ζ, using the iteration described below Theorem 30.

4. Compute the performance of each strategy using Theorems 25-27, as appropriate.

4.7.1 Escape time and quantizer bound

We compute the residence time through two methods, the analytical method based on Theorem 30 and the simulation. In addition, we compare the performance of coding strategies. Let us define the parameters as follow

- R_c, control weights in LQ output feedback control.
- $A - BK$, closed-loop pole of LQ.
- ζ, quantization bound.
- τ_a, mean escape time computed via Theorem 30.
- τ_{emp}, empirical mean escape time from simulation.
- J_I, J_{II}, corresponding performances for strategy I and II.

In the simulation for computing, τ_{emp}, we make the average over 20000 iterations of computing the very first time that the output signal jumps out of the quantizer bound for 5000 sample limit with the following parameter for the scalar system. Then we compare the performance of two different strategies with a fixed time $\tau = 1000$ and the parameters as follow

```
A = 0.9999; B = 1; C = 1; Q = 1;
R = 1; Qc = 1;
```

for 3-bit quantizer

R_c	A-BK	ζ	τ_a	τ_{emp}	J_I	J_{II}
1e5	0.9968	43.14	1000	2320	325	309
1e4	0.9900	24.67	1000	2194	104	101
1e3	0.9689	14.50	1000	1813	34.136	34.135
100	0.9049	9.15	1000	1317	11.81	12.43
10	0.7298	6.62	1000	1040	4.78	5.56
1	0.3819	5.68	1000	990	2.64	3.45
0.1	0.0839	5.49	998	977	2.10	2.92

for 2-bit quantizer

R_c	A-BK	ζ	τ_a	τ_{emp}	J_I	J_{II}
1e5	0.9968	48.14	1000	2354	474	315
1e4	0.9900	27.74	1000	2159	137	103
1e3	0.9689	16.44	1000	1780	42.22	35.04
100	0.9049	10.43	1000	1413	14.34	12.97
10	0.7298	7.54	1000	1213	5.94	5.91
1	0.3819	6.40	1000	1164	3.43	3.73
0.1	0.0839	6.12	998	1146	2.81	3.18

As we may conclude from the above example, the coding strategy is picked based on the nature of the controlled output signal. If the output signal has random or unpredictable nature, Figure 4.2, the coding has less benefits and we stick with Strategy I. In contrast, the coding strategy

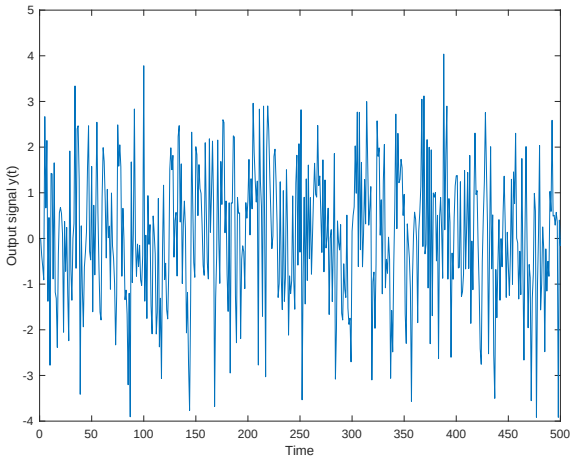

Figure 4.2. Controlled output signal y(t) with 3-bit coding and $R_c = 0.01$, corresponding to roughly minimum-variance control and hence to low amplitude, near-white y_t. Coding provides little benefit.

has advantages if the output controlled signal is more regulated or predictable such as Figure 4.3. In this case, as we have higher resolution or accuracy including a delay in updating the measurement, Strategy II outperforms Strategy I in which the measurement is updated each time but with less accuracy. When the control objective is minimum variance, the output signal resembles to a white noise signal and the quantization bound has smaller size, so the coding has no benefits. Once we move away from minimum variance control objective the output signal y_k has larger amplitude, furthermore the output signal y_k is correlated.

We may wrap up the following results from this example,

- If R_c is small, no benefit is obtained from coding.
- If bits, b, is large then coding has limited benefits.
- The quantizer bound, ζ, is smaller for the better regulated signal y_t.
- The controller with smaller R_c leads to smaller and whiter signal y_t.

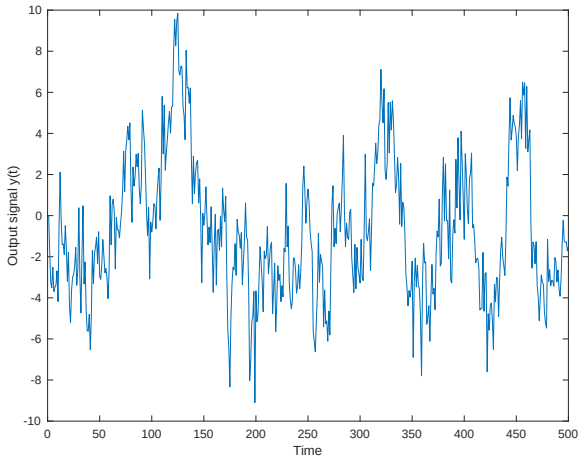

Figure 4.3. Controlled output signal y(t) with 3-bit coding and $R_c = 100$, corresponding to higher amplitude, correlated y_t. Coding provides tangible control benefit.

4.8 Conclusion

We have explored three very specific periodic coding strategies of the plant output signal and their effect on LQ performance subject to an expected escape time. The interaction between the control law and the estimation problem occurs through the selection of the upper bound, ζ, of the dithered quantizers. The general conclusion is that the more correlated is the controlled output, the more benefit is achieved by coding. So that minimum variance problems should exhibit less gain from coding than should those with heavier control penalty. The computational examples show that these coding schemes is of most value when the number of bits is small. These are generalizable conclusions to other more sophisticated codes and reflect observations in signal processing, but without the connection to a control objective.

The novelties of the approach lie in the treatment of the dithered quantizers and the introduction of the system escape time as a tool for analysis. The decomposition of the quantizer into two parts – infinite quantizer plus saturation – together with the escape time permits the consideration of linear controlled covariances and the distinct escape time analysis. These study of escape time is distinguished from other studies which seek to manage asymptotic or

infinite-horizon average properties.

4.9 Appendix

Proof for lemma 20:

Let us start with initial state estimate $\hat{x}_{0|-1}$ and covariance $\Sigma_{0|-1}$, the Kalman filter is calculated by:

At even time, $t = 2k$: (Low resolution measurement)

$$L_{2k} = \Sigma_{2k|2k-1}C^T(C\Sigma_{2k|2k-1}C^T + R + S_b)^{-1},$$
$$\hat{x}_{2k|2k} = \hat{x}_{2k|2k-1} + L_{2k}(p_{2k} - C\hat{x}_{2k|2k-1}),$$

At odd time, $t = 2k+1$: (High resolution measurement) We receive the less significant part of the quantized y_{2k} and construct the $2b$-bits measurement $z_{2b,2k+1} = z_{b,2k} \oplus_{2b,2k}$ through concatenation,

$$p_{2k+1} \leftarrow p_{2k} + 2^{-b}m_{2k+1} - d_{2k}^{2b}$$
$$L_{2k+1} = \Sigma_{2k|2k-1}C^T(C\Sigma_{2k|2k-1}C^T + R + S_{2b})^{-1},$$
$$\hat{x}_{2k+1|2k+1} = A\hat{x}_{2k|2k+1} + Bu_{2k},$$
$$\hat{x}_{2k+1|2k+1} = A\left(\hat{x}_{2k|2k-1} + L_{2k+1}(p_{2k+1} - C\hat{x}_{2k|2k-1})\right) - BK\hat{x}_{2k|2k},$$
$$\hat{x}_{2k+2|2k+1} = (A - BK)\hat{x}_{2k+1|2k+1},$$
$$\Sigma_{2k+2|2k+1} = A^2\Sigma_{2k|2k-1}A^{2^T} - A^2L_{2k+1}C\Sigma_{2k|2k-1}A^{2^T} + AQA^T + Q.$$

Proof for Lemma 21:

Let us start with initial state estimate $\hat{x}_{0|-1}$ and covariance $\Sigma_{0|-1}$, at $t = 0$, the Kalman filter is calculated by:

At even time, $t = 2k$,

$$L_{2k} = \Sigma_{2k|2k-1}C^T(C\Sigma_{2k|2k-1}C^T + R + S_b)^{-1},$$
$$\hat{x}_{2k|2k} = \hat{x}_{2k|2k-1} + L_{2k}(p_{2k} - C\hat{x}_{2k|2k-1}),$$

At odd times, $t = 2k+1$,

$$\begin{aligned}
p'_{2k} &= p_{2k} + 2^{-b}\mathrm{MSB}_r(m_{2k+1}) - d_{2k}^{b+r},\\
L'_{2k} &= \Sigma_{2k|2k-1}C^T(C\Sigma_{2k|2k-1}C^T + R + S_{b+r})^{-1},\\
\hat{x}'_{2k+1|2k} &= A\hat{x}_{2k|2k-1} + AL'_{2k}(p'_{2k} - C\hat{x}_{2k|2k-1})\\
&\qquad - BK\hat{x}_{2k|2k},\\
\Sigma'_{2k+1|2k} &= A\Sigma_{2k|2k-1}A^T - A\Sigma_{2k|2k-1}C^T \times\\
&\quad (C\Sigma_{2k|2k-1}C^T + R + S_{b+r})^{-1}C\Sigma_{2k|2k-1}A^T + Q,\\
p_{2k+1} &= \mathrm{LSB}_{b-r}(m_{2k+1}) - d_{2k+1}^{b-r},\\
L_{2k+1} &= \Sigma'_{2k+1|2k}C^T(C\Sigma'_{2k+1|2k}C^T + R + S_{b-r})^{-1},\\
\hat{x}_{2k+1|2k+1} &= \hat{x}'_{2k+1|2k} + L_{2k+1}(p_{2k+1} - C\hat{x}'_{2k+1|2k}),\\
\hat{x}_{2k+2|2k+1} &= A\hat{x}_{2k+1|2k+1} + Bu_{2k+1},\\
\hat{x}_{2k+2|2k+1} &= (A - BK)\hat{x}_{2k+1|2k+1},\\
\Sigma_{2k+2|2k+1} &= A\Sigma'_{2k+1|2k}A^T - A\Sigma'_{2k+1|2k}C^T \times\\
&\quad (C\Sigma'_{2k+1|2k}CT + R + S_{b-r})^{-1}\Sigma'_{2k+1|2k}A^T + Q.
\end{aligned}$$

Proof for Corollary 22:

Can be found at [29] but the difference is quantization noise S_b is added to the measurement noise, and it is replaced by $R + S_b$ in all calculation.

Proof for Corollary 23:

Let us start with $\Sigma \triangleq \Sigma_{2k|2k-1}$

(i) Low resolution measurement

$$\Sigma_{2k|2k} = \Sigma - \Sigma C^T (C\Sigma C^T + R + S_b)^{-1} C\Sigma \tag{4.19}$$

(ii) High resolution measurement

$$\Sigma_{2k|2k+1} = \Sigma - \Sigma C^T (C\Sigma C^T + R + S_{2b})^{-1} C\Sigma \tag{4.20}$$

(iii) Time $2k+1$ filtered measurement

$$\Sigma_{2k+1|2k+1} = A\Sigma_{2k|2k+1}A^T + Q.$$

(iv) Time $2k+2$ prediction and let $k \to \infty$

$$\begin{aligned}\Sigma_{2k+2|2k+1} &= A\Sigma_{2k+1|2k+1}A^T + Q, \\ &= A^2\Sigma_{2k|2k+1}A^{2^T} + AQA^T + Q,\end{aligned}$$

$$\begin{aligned}\Sigma_{II}^{p,\infty} = A^2\Sigma_{II}^{p,\infty}A^{2^T} - A^2\Sigma_{II}^{p,\infty}C^T \times \\ (C\Sigma_{II}^{p,\infty}C^T + R + S_{2b})^{-1} C\Sigma_{II}^{p,\infty}A^{2^T} + \\ AQA^T + Q.\end{aligned}$$

where

$$\lim_{k\to\infty} \Sigma_{2k|2k-1} = \Sigma_{II}^{p,\infty} = \texttt{dare}\left(A^{2^T}, C^T, AQA^T + Q, R + S_{2b}\right).$$

(v) Substitute $\Sigma_{II}^{p,\infty}$ into (4.19) and (4.20)

$$\begin{aligned}
\Sigma_{II}^{p,\infty} &= \texttt{dare}\left(A^{2^T}, C^T, AQA^T + Q, R + S_{2b}\right), \\
\Sigma_{II_{\text{even}}}^{\infty} &= \Sigma_{II}^{p,\infty} - \Sigma_{II}^{p,\infty} C^T (C\Sigma_{II}^{p,\infty} C^T + R + S_b)^{-1} C\Sigma_{II}^{p,\infty}, \\
\Sigma_{II_{\text{odd}}}^{\infty} &= \Sigma_{II}^{p,\infty} - \Sigma_{II}^{p,\infty} C^T (C\Sigma_{II}^{p,\infty} C^T + R + S_{2b})^{-1} C\Sigma_{II}^{p,\infty}.
\end{aligned}$$

Proof for Corollary 24:

The period-two update consists of two pieces starting from the same initial data, $(\hat{x}_{2k|2k-1}, \Sigma = \Sigma_{2k|2k-1})$.

Even times – No need to keep track of this in the computation of the covariance $\Sigma_{2k+1|2k+1}$ since this is calculated based only on

$$\begin{aligned}
z'_{2k} &= Cx_{2k} + v_{2k} + q_{b+r,2k}, \\
z_{2k+1} &= Cx_{2k+1} + v_{2k+1} + q_{b-r,2k+1}.
\end{aligned}$$

It is, however, important for the Lyapunov computation.

Odd times – We skip over the even step and use both z'_{2k} and z_{2k+1} to update $\hat{x}_{2k|2k-1}$. Start as usual.

$$\begin{aligned}
x_{2k+1} &= Ax_{2k} + Bu_{2k} + w_{2k}, \\
p_{2k+1} &= Ax_{2k} + w_{2k}, \\
z'_{2k} &= Cx_{2k} + v_{2k} + q_{b+r,2k}, \\
z_{2k+1} &= CAx_{2k} + CBu_{2k} + Cw_{2k} + v_{2k+1} + q_{b-r,2k+1}, \\
\zeta_{2k+1} &= CAx_{2k} + Cw_{2k} + v_{2k+1} + q_{b-r,2k+1}
\end{aligned}$$

where denote $\zeta_{2k+1} = z_{2k+1} - CBu_{2k}$ and calculate joint conditional density,

$$\text{pdf}\left(\left.\begin{bmatrix} p_{2k+1} \\ z'_{2k} \\ \zeta_{2k+1} \end{bmatrix}\right| \mathscr{Z}^{2k-1}\right) = \mathscr{N}\left(\begin{bmatrix} A\hat{x}_{2k|2k-1} \\ C\hat{x}_{2k|2k-1} \\ CA\hat{x}_{2k|2k-1} \end{bmatrix}, \mathscr{M}\right),$$

$$\mathscr{M} = \begin{bmatrix} A\Sigma A^T + Q & A\Sigma C^T & A\Sigma A^T C^T + QC^T \\ C\Sigma A^T & C\Sigma C^T + R + S_{b+r} & C\Sigma A^T C^T \\ CA\Sigma A^T + CQ & CA\Sigma C^T & CA\Sigma A^T C^T + CQC^T + R + S_{b-r} \end{bmatrix},$$

hence,

$$\begin{aligned} \text{cov}(x_{2k+1}|\mathscr{Z}^{2k+1}) = A\Sigma A^T + Q - \begin{bmatrix} A\Sigma C^T & A\Sigma A^T C^T + QC^T \end{bmatrix} \times \\ \begin{bmatrix} C\Sigma C^T + R + S_{b+r} & C\Sigma A^T C^T \\ CA\Sigma C^T & CA\Sigma A^T C^T + CQC^T + R + S_{b-r} \end{bmatrix}^{-1} \times \\ \begin{bmatrix} C\Sigma A^T \\ CA\Sigma A^T + CQ \end{bmatrix}, \end{aligned}$$

by taking limits as

$$\lim_{k\to\infty} \Sigma_{2k|2k-1} = \Sigma_{2k+2|2k+1} = \Sigma,$$

$$\text{cov}(x_{2k+2}|\mathscr{Z}^{2k+1}) = A \times \text{cov}(x_{2k+1}|\mathscr{Z}^{2k+1}) \times A^T + Q,$$

$$\Sigma = A^{2}\Sigma A^{2^T} + AQA^T + Q - \begin{bmatrix} A^{2}\Sigma C^T & A^{2}\Sigma A^T C^T + AQC^T \end{bmatrix} \times \begin{bmatrix} C\Sigma C^T + R + S_{b+r} & C\Sigma A^T C^T \\ CA\Sigma C^T & CA\Sigma A^T C^T + CQC^T + R + S_{b-r} \end{bmatrix}^{-1} \times \begin{bmatrix} C\Sigma A^{2^T} \\ CA\Sigma A^{2^T} + CQA^T \end{bmatrix},$$

$$= A^{2}\Sigma A^{2^T} + AQA^T + Q - \left(A^{2}\Sigma \begin{bmatrix} C^T & A^T C^T \end{bmatrix} + \begin{bmatrix} 0 & AQC^T \end{bmatrix} \right) \times \left(\begin{bmatrix} C \\ CA \end{bmatrix} \Sigma \begin{bmatrix} C^T & A^T C^T \end{bmatrix} + \begin{bmatrix} R + S_{b+r} & 0 \\ 0 & CQC^T + R + S_{b-r} \end{bmatrix} \right)^{-1} \times \left(\begin{bmatrix} C \\ CA \end{bmatrix} \Sigma A^{2^T} + \begin{bmatrix} 0 \\ CQA^T \end{bmatrix} \right).$$

and we use DARE to calculate,

$$\Sigma_{III}^{p,\infty} = \texttt{dare}\left(A^{2^T}, \begin{bmatrix} C^T & A^T C^T \end{bmatrix}, AQA^T + Q, \mathscr{W}, \begin{bmatrix} 0 & AQC^T \end{bmatrix}, \texttt{eye}(n) \right),$$

and similar to proof of Corollary 23

$$\Sigma_{III_{\text{even}}}^{\infty} = \Sigma_{III}^{p,\infty} - \Sigma_{III}^{p,\infty} C^T \left(C\Sigma_{III}^{p,\infty} C^T + R + S_{b+r} \right)^{-1} C\Sigma_{III}^{p,\infty},$$
$$\Sigma_{III_{\text{odd}}}^{\infty} = \Sigma_{III}^{p,\infty} - \Sigma_{III}^{p,\infty} C^T \left(C\Sigma_{III}^{p,\infty} C^T + R + S_{b-r} \right)^{-1} C\Sigma_{III}^{p,\infty},$$

where,

$$\mathscr{W} = \begin{bmatrix} R + S_{b+r} & 0 \\ 0 & CQC^T + R + S_{b-r} \end{bmatrix}.$$

Proof for Theorem 25:

Truncate every sample to b bits, transmit

$$z_t = Cx_t + v_t + q_{b,t}.$$

Kalman filter is stationary and satisfies

$$\begin{aligned}\Sigma &= \texttt{dare}\left(A^T, C^T, Q, R + S_b\right), \\ L &= \Sigma - \Sigma C^T \left(C\Sigma C^T + R + S_b\right).\end{aligned}$$

Closed-loop equations

$$
\begin{aligned}
x_{t+1} &= Ax_t - BK\hat{x}_{t|t} + w_t, \\
\hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + L(Cx_{t+1} + v_{t+1} + q_{b,t+1} \\
&\qquad - C\hat{x}_{t+1|t}), \\
&= (A\hat{x}_{t|t} - BK\hat{x}_{t|t}) \\
&\quad + L\left[C(Ax_t - BK\hat{x}_{t|t} + w_t)\right] \\
&\quad + L\left[v_{t+1} + q_{b,t+1} - C(A-BK)\hat{x}_{t|t}\right], \\
&= [(I-LC)(A-BK) - LCBK]\hat{x}_{t|t} \\
&\quad + LCAx_t + LCw_t \\
&\quad + Lv_{t+1} + Lq_{b,t+1}, \\
&= LCAx_t + [(I-LC)A - BK]\hat{x}_{t|t} \\
&\quad + LCw_t + Lv_{t+1} + Lq_{b,t+1},
\end{aligned}
$$

$$
\begin{bmatrix} x_{t+1} \\ \hat{x}_{t+1|t+1} \end{bmatrix} = \begin{bmatrix} A & -BK \\ LCA & (I-LC)A - BK \end{bmatrix} \begin{bmatrix} x_t \\ \hat{x}_{t|t} \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ LC & L & L \end{bmatrix} \begin{bmatrix} w_t \\ v_{t+1} \\ q_{b,t+1} \end{bmatrix}.
$$

Let us denote

$$
\mathscr{A} = \begin{bmatrix} A & -BK \\ LCA & (I-LC)A - BK \end{bmatrix}, \quad \mathscr{B} = \begin{bmatrix} I & 0 & 0 \\ LC & L & L \end{bmatrix},
$$

$$\mathscr{Q} = \begin{bmatrix} Q & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & S_b \end{bmatrix}, \ \Psi_I = \texttt{dlyap}\left(\mathscr{M}_1, \mathscr{N}_1 \mathscr{P}_1 \mathscr{N}_1^T\right),$$

hence the performance calculation,

$$J_I = \text{trace}[Q_c \Psi_I(1,1)] + \text{trace}[K^T R_c K \Psi_I(2,2)],$$

where,

$$\Psi_I = \begin{bmatrix} E(x_k x_k^T) & E(x_k \hat{x}_{k|k}^T) \\ E(\hat{x}_{k|k} x_k^T) & E(\hat{x}_{k|k} \hat{x}_{k|k}^T) \end{bmatrix}.$$

Proof for Theorem 26:

Truncate y_t to b bits at even times t and then to $2b$ bits at odd times t. The quantization variances S_b and S_{2b} respectively.

Start with x_{2k-1}, $\hat{x}_{2k-1|2k-1}$ and $\Sigma_{2k|2k-1}$. Compute

$$L_{2k} = \Sigma_{2k|2k-1} C^T (C\Sigma_{2k|2k-1} C^T + R + S_b)^{-1},$$
$$L_{2k+1} = \Sigma_{2k|2k-1} C^T (C\Sigma_{2k|2k-1} C^T + R + S_{2b})^{-1}.$$

State and predictor update

$$x_{2k} = A x_{2k-1} - BK\hat{x}_{2k-1|2k-1} + w_{2k-1}.$$
$$\hat{x}_{2k|2k-1} = (A - BK)\hat{x}_{2k-1|2k-1},$$

so rearrange these equations,

$$\begin{bmatrix} \hat{x}_{2k|2k-1} \\ x_{2k} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & A-BK \\ 0 & 0 & A & -BK \end{bmatrix} \begin{bmatrix} x_{2k-2} \\ \hat{x}_{2k-2|2k-2} \\ x_{2k-1} \\ \hat{x}_{2k-1|2k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} w_{2k-1}$$

Filter update with low resolution measurement, $z_{b,2k}$.

$$\hat{x}_{2k|2k} = \hat{x}_{2k|2k-1} + L_{2k}(z_{b,2k} - C\hat{x}_{2k|2k-1}),$$

rearrange the equation in matrix form,

$$\begin{bmatrix} \hat{x}_{2k|2k-1} \\ x_{2k} \\ \hat{x}_{2k|2k} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \\ (I-L_{2k}C) & L_{2k}C \end{bmatrix} \begin{bmatrix} \hat{x}_{2k|2k-1} \\ x_{2k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ L_{2k} \end{bmatrix} v_{2k} + \begin{bmatrix} 0 \\ 0 \\ L_{2k} \end{bmatrix} q_{2k}.$$

Filter update with high resolution measurement,

$$z_{2k+1} = z_{b,2k} \oplus z_{2b,2k} = Cx_{2k} + v_{2k} + q_{2k+1},$$

$$\hat{x}_{2k|2k+1} = \hat{x}_{2k|2k-1} + L_{2k+1}(z_{2k+1} - C\hat{x}_{2k|2k-1}),$$

$$\begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k} \\ \hat{x}_{2k|2k+1} \end{bmatrix} = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ (I - L_{2k+1}C) & L_{2k+1}C & 0 \end{bmatrix} \begin{bmatrix} \hat{x}_{2k|2k-1} \\ x_{2k} \\ \hat{x}_{2k|2k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ L_{2k+1} \end{bmatrix} v_{2k} + \begin{bmatrix} 0 \\ 0 \\ L_{2k+1} \end{bmatrix} q_{2k+1}.$$

State and state estimate update.

$$x_{2k+1} = Ax_{2k} - BK\hat{x}_{2k|2k} + w_{2k},$$
$$\hat{x}_{2k+1|2k+1} = A\hat{x}_{2k|2k+1} - BK\hat{x}_{2k|2k}.$$

denote,

$$\begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k} \\ x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ A & -BK & 0 \\ 0 & -BK & A \end{bmatrix} \begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k} \\ \hat{x}_{2k|2k+1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \end{bmatrix} w_{2k}.$$

Now define

$$\lim_{k\to\infty} L_{2k} = L_{\text{even}} = \Sigma_{II}^{p,\infty} C^T \left(C\Sigma_{II}^{p,\infty} C^T + R + S_b \right)^{-1},$$
$$\lim_{k\to\infty} L_{2k+1} = L_{\text{odd}} = \Sigma_{II}^{p,\infty} C^T (C\Sigma_{II}^{p,\infty} C^T + R + S_{2b})^{-1},$$

$$F_1 = \begin{bmatrix} 0 & 0 & 0 & A-BK \\ 0 & 0 & A & -BK \end{bmatrix}, \quad F_2 = \begin{bmatrix} I & 0 \\ 0 & I \\ (I - L_{\text{even}}C) & L_{\text{even}}C \end{bmatrix},$$

$$F_3 = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ (I - L_{\text{odd}}C) & L_{\text{odd}}C & 0 \end{bmatrix}, \; F_4 = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ A & -BK & 0 \\ 0 & -BK & A \end{bmatrix},$$

$$G_1 = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 \\ 0 \\ L_{\text{even}} \end{bmatrix}, \quad G_3 = \begin{bmatrix} 0 \\ 0 \\ L_{\text{odd}} \end{bmatrix}, \quad G_4 = \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \end{bmatrix}$$

Then, the two-step update is described by the recursion

$$\begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k} \\ x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \end{bmatrix} = \mathscr{M}_2 \begin{bmatrix} x_{2k-2} \\ \hat{x}_{2k-2|2k-2} \\ x_{2k-1} \\ \hat{x}_{2k-1|2k-1} \end{bmatrix} + \mathscr{N}_2 \begin{bmatrix} w_{2k-1} \\ w_{2k} \\ \begin{bmatrix} v_{2k} + q_{2k} \\ v_{2k} + q_{2k+1} \end{bmatrix} \end{bmatrix},$$

with

$$\mathscr{M}_2 = F_4F_3F_2F_1, \quad \mathscr{N}_2 = \begin{bmatrix} F_4F_3F_2G_1 & G_4 & F_4F_3G_2 & F_4G_3 \end{bmatrix}.$$

Whence,

$$\Psi_{II} = E\left(\begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k} \\ x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \end{bmatrix} \begin{bmatrix} x_{2k}^T & \hat{x}_{2k|2k}^T & x_{2k+1}^T & \hat{x}_{2k+1|2k+1}^T \end{bmatrix}\right),$$

$$\Psi_{II} = \texttt{dlyap}\left(\mathscr{M}_2, \mathscr{N}_2 \begin{bmatrix} Q & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & R+S_b & R+S_{2b} \\ 0 & 0 & R+S_{2b} & R+S_{2b} \end{bmatrix} \mathscr{N}_1^T\right),$$

and

$$\begin{aligned} J_{II} = &\frac{1}{2}\text{trace}\left\{Q_c[\Psi_{II}(1,1)+\Psi_{II}(3,3)]\right\}+ \\ &\frac{1}{2}\text{trace}\left\{K^T R_c K[\Psi_{II}(2,2)+\Psi_{II}(4,4)]\right\}, \end{aligned}$$

Proof for Theorem 27:

From $\Sigma = \Sigma_{2k|2k-1}$, compute the filter gains

$$\begin{aligned} L_{2k} &= \Sigma_{2k|2k-1}C^T(C\Sigma_{2k|2k-1}C^T + R + S_b)^{-1}, \\ L'_{2k} &= \Sigma_{2k|2k-1}C^T(C\Sigma_{2k|2k-1}C^T + R + S_{b+r})^{-1}, \\ \Sigma_{2k+1|2k} &= A\Sigma'_{2k|2k}A^T + Q, \\ &= A\Sigma A^T - A\Sigma C^T(C\Sigma C^T + R + S_{b+r})^{-1}C\Sigma A^T + Q, \\ L_{2k+1} &= \Sigma_{2k+1|2k}C^T(C\Sigma_{2k+1|2k}C^T + S_{b-r})^{-1}. \end{aligned}$$

denote the quantization noises $S_b = \frac{\zeta^2}{3\times 2^{2b}}$

$$S_{b+r} = \frac{\zeta^2}{3\times 2^{2(b+r)}},\ \ S_{b-r} = \frac{\zeta^2}{3\times 2^{2(b-r)}}.$$

Let us start with $\begin{bmatrix} x_{2k-1} \\ \hat{x}_{2k-1|2k-1} \\ x_{2k-2} \\ \hat{x}^1_{2k-2|2k-2} \end{bmatrix}$ and then,

$$\begin{aligned}
x_{2k} &= Ax_{2k-1} - BK\hat{x}_{2k-1|2k-1} + w_{2k-1},\\
\hat{x}_{2k|2k-1} &= A\hat{x}_{2k-1|2k-1} - BK\hat{x}_{2k-1|2k-1},\\
&= (A-BK)\hat{x}_{2k-1|2k-1},\\
\hat{x}^1_{2k|2k} &= (I-L_{2k}C)\hat{x}_{2k|2k-1} + L_{2k}z^1_{2k},\\
&= (I-L_{2k}C)\hat{x}_{2k|2k-1} + L_{2k}Cx_{2k}\\
&\quad + L_{2k}v_{2k} + L_{2k}q_{b,2k},\\
\hat{x}'_{2k|2k} &= (I-L'_{2k}C)\hat{x}_{2k|2k-1} + L'_{2k}Cx_{2k}\\
&\quad + L'_{2k}v_{2k} + L'_{2k}q_{b+r,2k},\\
x_{2k+1} &= Ax_{2k} - BK\hat{x}^1_{2k|2k} + w_{2k},\\
\hat{x}_{2k+1|2k} &= A\hat{x}'_{2k|2k} - BK\hat{x}^1_{2k|2k},\\
\hat{x}_{2k+1|2k+1} &= (I-L_{2k+1}C)\hat{x}_{2k+1|2k} + L_{2k+1}z_{2k+1},\\
&= (I-L_{2k+1}C)A\hat{x}'_{2k|2k}\\
&\quad - (I-L_{2k+1}C)BK\hat{x}^1_{2k|2k}\\
&\quad + L_{2k+1}Cx_{2k+1} + L_{2k+1}v_{2k+1}\\
&\quad + L_{2k+1}q_{b-r,2k+1}.
\end{aligned}$$

The short sequence.

$$\begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k-1} \end{bmatrix} = \begin{bmatrix} A & -BK & 0 & 0 \\ 0 & A-BK & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{2k-1} \\ \hat{x}_{2k-1|2k-1} \\ x_{2k-2} \\ \hat{x}^1_{2k-2|2k-2} \end{bmatrix} + \begin{bmatrix} I \\ 0 \end{bmatrix} w_{2k-1},$$

$$\begin{bmatrix} x_{2k} \\ \hat{x}^1_{2k|2k} \\ \hat{x}'_{2k|2k} \end{bmatrix} = \begin{bmatrix} I & 0 \\ L_{2k}C & (I-L_{2k}C) \\ L'_{2k}C & (I-L'_{2k}C) \end{bmatrix} \begin{bmatrix} x_{2k} \\ \hat{x}_{2k|2k-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ L_{2k} & 0 \\ 0 & L'_{2k} \end{bmatrix} \begin{bmatrix} v_{2k}+q_{b,2k} \\ v_{2k}+q_{b+r,2k} \end{bmatrix},$$

$$\begin{bmatrix} x_{2k+1} \\ \hat{x}'_{2k|2k} \\ x_{2k} \\ \hat{x}^1_{2k|2k} \end{bmatrix} = \begin{bmatrix} A & -BK & 0 \\ 0 & 0 & I \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} x_{2k} \\ \hat{x}^1_{2k|2k} \\ \hat{x}'_{2k|2k} \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix} w_{2k},$$

$$\begin{bmatrix} x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \\ x_{2k} \\ \hat{x}^1_{2k|2k} \end{bmatrix} =$$

$$\begin{bmatrix} I & 0 & 0 & 0 \\ L_{2k+1}C & (I-L_{2k+1}C)A & 0 & -(I-L_{2k+1}C)BK \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix} \times \begin{bmatrix} x_{2k+1} \\ \hat{x}'_{2k|2k} \\ x_{2k} \\ \hat{x}^1_{2k|2k} \end{bmatrix} + \begin{bmatrix} 0 \\ L_{2k+1} \\ 0 \\ 0 \end{bmatrix} (v_{2k+1}+q_{b-r,2k+1})$$

denote

$$\lim_{k\to\infty} L_{2k} = L_{\text{even}} = \Sigma^{p,\infty} C^T \left(C\Sigma^{p,\infty} C^T + R + S_b\right)^{-1},$$

$$\lim_{k\to\infty} L'_{2k} = L_{\text{odd1}} = \Sigma^{p,\infty} C^T (C\Sigma^{p,\infty} C^T + R + S_{b+r})^{-1},$$

$$\lim_{k\to\infty} L_{2k+1} = L_{\text{odd2}} = \Sigma^{p,\infty} C^T (C\Sigma^{p,\infty} C^T + R + S_{b-r})^{-1},$$

$$F_1 = \begin{bmatrix} A & -BK & 0 & 0 \\ 0 & A-BK & 0 & 0 \end{bmatrix}, \; G_1 = \begin{bmatrix} I \\ 0 \end{bmatrix},$$

$$F_2 = \begin{bmatrix} I & 0 \\ L_{\text{even}}C & (I - L_{\text{even}}C) \\ L_{\text{odd1}}C & (I - L_{\text{odd1}}C) \end{bmatrix}, G_2 = \begin{bmatrix} 0 & 0 \\ L_{\text{even}} & 0 \\ 0 & L_{\text{odd1}} \end{bmatrix},$$

$$F_3 = \begin{bmatrix} A & -BK & 0 \\ 0 & 0 & I \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix}, \; G_3 = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix}, \; G_4 = \begin{bmatrix} 0 \\ L_{\text{odd2}} \\ 0 \\ 0 \end{bmatrix},$$

$$F_4 = \begin{bmatrix} I & 0 & 0 & 0 \\ L_{\text{odd2}}C & (I - L_{\text{odd2}}C)A & 0 & -(I - L_{\text{odd2}}C)BK \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}.$$

Then, the two-step update is described by

$$\begin{bmatrix} x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \\ x_{2k} \\ \hat{x}^1_{2k|2k} \end{bmatrix} = \mathscr{M}_3 \begin{bmatrix} x_{2k-1} \\ \hat{x}_{2k-1|2k-1} \\ x_{2k-1} \\ \hat{x}^1_{2k-2|2k-2} \end{bmatrix} + \mathscr{N}_3 \begin{bmatrix} w_{2k-1} \\ w_{2k} \\ \begin{bmatrix} v_{2k} + q_{b,2k} \\ v_{2k} + q_{b+r,2k} \end{bmatrix} \\ v_{2k+1} + q_{b-r,2k+1} \end{bmatrix},$$

with

$$\begin{aligned} \mathscr{M}_3 &= F_4 F_3 F_2 F_1, \\ \mathscr{N}_3 &= \begin{bmatrix} F_4 F_3 F_2 G_1 & F_4 G_3 & F_4 F_3 G_2 & G_4 \end{bmatrix}, \end{aligned}$$

$$\Psi_{III} = E\left(\begin{bmatrix} x_{2k+1} \\ \hat{x}_{2k+1|2k+1} \\ x_{2k} \\ \hat{x}^1_{2k|2k} \end{bmatrix} \begin{bmatrix} x^T_{2k+1} & \hat{x}^T_{2k+1|2k+1} & x^T_{2k} & \hat{x}^{1^T}_{2k|2k} \end{bmatrix} \right),$$

$$\Psi_{III} = \texttt{dlyap}\left(\mathscr{M}_3, \mathscr{N}_3 \mathscr{P}_3 {\mathscr{N}_3}^T \right),$$

$$\mathscr{P}_3 = \begin{bmatrix} Q & 0 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 & 0 \\ 0 & 0 & R+S_b & R+S_{b+r} & 0 \\ 0 & 0 & R+S_{b+r} & R+S_{b+r} & 0 \\ 0 & 0 & 0 & 0 & R+S_{b-r} \end{bmatrix},$$

$$J_{III} = \frac{1}{2}\text{trace}\left\{Q_c[\Psi_{III}(1,1) + \Psi_{III}(3,3)]\right\} + \frac{1}{2}\text{trace}\left\{K^T R_c K[\Psi_{III}(2,2) + \Psi_{III}(4,4)]\right\}.$$

Proof for Theorem 30:

Suppose $\Pr(|z_k| > \zeta) = \beta$, then $\Pr(|z_k| < \zeta) = 1 - \beta$, for $t = 1, 2, ...$ and assuming the events to be independent,

$$\Pr[(|z_1| < \zeta) \cap (|z_2| < \zeta) ... \cap (|z_N| < \zeta)] = (1-\beta)^N.$$

The probability that the process escapes at time T is computed as $(1-\beta)^{T-1}\beta$ and the expected time is

$$\begin{aligned}
\mathrm{E}[\tau_{esc}] &= \beta + 2(1-\beta)\beta + 3(1-\beta)^2\beta + 4(1-\beta)^3\beta + ..., \\
&= \beta[1 + 2(1-\beta) + 3(1-\beta)^2 + 4(1-\beta)^3 + ...], \\
&= \beta\frac{d}{d\beta}[\frac{-1}{1-(1-\beta)}], \\
&= \beta\frac{d}{d\beta}\left[\frac{-1}{\beta}\right], \\
&= \frac{1}{\beta}.
\end{aligned}$$

ACKNOWLEDGEMENTS

I would like to acknowledge Chapter 3 is from the paper,“Predictive coding and control”, IEEE Transaction On Control of Network Systems, , Chun-Chia Huang, Behrooz Amini and Robert R. Bitmead 2018. Chapter 4 is from the paper,“LQG Control Performance with Low Bitrate Periodic Coding,” submitted to IEEE Transaction on Control of Network Systems, Behrooz Amini, Robert R. Bitmead 2020.

Chapter 5

Conclusions and future directions

A comprehensive survey of prior research in quantization process is reviewed with the goal to apply in control systems, and various conditions are studied to catch the proper stipulations to ameliorate the side effects of quantization error. Eventually we refine the results of previous achievements to treat the quantization error as *white noise.* This results play a conspicuous role in the rest of our study surely. We have introduced and applied subtractive dither quantization in a network control systems.

We have presented a complete assessment of predictive coding when used as a part of network control system. We have shown that feedback control based on transmission of the innovations sequence can not stabilize an unstable system because the unstable mode is not detectable.

We have investigated the application of Bayesian filter for control systems based on a predictive coding method. The predictive coding conveys efficiency in the use of the channel bits and capacity. This leads to develop state estimation at the receiver and, in turn, to a more accurate state density at the receiver. We assessed the LQG control performance for the controlled closed loop feedback and the performance of three controllers was computed using LQ-optimal feedback gain and the various conditional mean state estimates. We have concluded the controller is based on filtered state estimate from quantized innovation outperforms the controller is produced from the filtered state estimate. Clearly the worst performance pertains to the controller produced from the quantized output. Hence we have shown the predictive coding has a significant advantage in

network control system and it could be very useful. In particular geographically distributed large scale network control systems. Of course this requires to be balanced against the computational cost of operation Bayesian filter at the receiver.

We have explored very specific periodic coding strategies of the plant output signal and their impact on linear quadratic performance subject to an expected escape time. We realized the more correlated is the controlled output, the more benefit is obtained by coding. The conclusion is that the coding is of most value when the number of bits is small. Through some basic coding strategies, we have shown the control objective function has a key role to play in the efficacy of coding depending on the nature of the controlled output signal.

The novel commodity of this study traced back to treat the quantization's saturation via introducing the escape time first and assess the LQG performance over that time. We have decomposed the quantizer in two stages as infinite levels quantizer and saturation. So it offers us to investigate the linear controlled covariances and escape time at once. The study of the escape time is distinguished from other studies which seek to focus on infinite-horizon average or to manage asymptotic properties.

The future and subsequent directions could be: we would introduce more complicate communication channel and the second channel from the controller to the plant. Incorporation of further channel defects such as packet-drop, delays are simple extensions of the Bayesian filter, additive noise and more complicated channel codings. The coding methods to be studies could cover, not just more sophisticated codes tuned to the control output signal properties, but error-correcting codes when the channel introduces errors. From a practical point, when we implement Bayesian filter, we have chosen a stationary problem to be able to compute on a static grid, while more generally Bayesian filter is realized through Particle filters.

Generally the innovations sequence no longer need to be white. Even in case of linear non-Gaussian, it is uncorrelated but not necessarily white. In addition we can extend the computation with fully nonlinear, time-varying state and measurement equations.

Bibliography

[1] Renwick E. Curry. *Estimation and Control with Quantized Measurements*, volume 60 of *Research Monograph*. MIT Press, Cambridge MA, 1970.

[2] Robert M Gray and Thomas G Stockham. Dithered quantizers. *IEEE Transactions on Information Theory*, 39(3):805–812, 1993.

[3] B. Widrow and I. Kollár. *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*. Cambridge Univ. Press, 2008.

[4] Anekal Sripad and Donald Snyder. A necessary and sufficient condition for quantization errors to be uniform and white. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(5):442–448, 1977.

[5] Leonard Schuchman. Dither signals and their effect on quantization noise. *IEEE Transactions on Communication Technology*, 12(4):162–165, 1964.

[6] A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, New York, NY, 1992.

[7] Telecommunication Standardization Sector. *7 kHz audio-coding within 64 kbit/s*. International Telecommunication Union, September 2012.

[8] H.M. Jones, R.R. Bitmead, and S. Crisafulli. Feedback control applied to speech coding. In *IEEE Conference on Decision and Control*, pages 1881–1885, Kobe, Japan, 1996.

[9] G. Nair and R.J. Evans. Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM Journal on Control and Optimization*, 43(2):413–436, 2004.

[10] K. You, L. Xie, S. Sun, and W. Xiao. Multiple-level quantized innovation Kalman filter. *Proc. 17th International Federation of Automatic Control World Congress, Seoul, Korea*, 2008.

[11] A. Ribeiro, G.B. Giannakis, and S.I. Roumeliotis. SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations. *IEEE Transactions on Signal Processing*, 54:4782–4795, 2006.

[12] R. Sukhavasi and B. Hassibi. The Kalman-like particle filter: Optimal estimation with quantized innovations/measurements. *IEEE Transactions on Signal Processing*, 61(1):131–136, January 2013.

[13] V.S. Borkar and S.K. Mitter. LQG control with communication constraints. In A. Paulraj, V. Roychowdhury, and C.D. Schaper, editors, *Communication, Computation, Control and Signal Processing*, chapter 21, pages 365–373. Kluwer Academic Publishers, New York NY, 1997.

[14] D.F. Delchamps. Extracting state information from quantized output record. *Systems & Control Letters*, 13:365–372, 1989.

[15] Wing Shing Wong and R.W. Brockett. Systems with finite communication bandwidth constraints. i. state estimation problems. *IEEE Transactions on Automatic Control*, 42(9):1294 –1299, September 1997.

[16] S-I Azuma and T. Sugie. Optimal dynamic quantizers for discrete-valued input control. *Automatica*, 44:396–406, 2008.

[17] H. Okajima, K. Sawada, and M. Matsunaga. Dynamic quantizer design under communication rate constraints. *IEEE Transactions on Automatic Control*, 61(10):3190–3196, 2016.

[18] T.R. Fischer. Optimal quantized control. *IEEE Trans Automatic Control*, 27(4):996–998, 1982.

[19] Minyue Fu. Lack of separation principle for quantized Linear Quadratic Gaussian control. *IEEE Trasactions on Automatic Control*, 57(9):2385–2390, 2012.

[20] Serdar Yüksel. Jointly optimal LQG quantization and control policies for multi-dimensional systems. *IEEE Transactions on Automatic Control*, 59(6):1612–1617, 2014.

[21] Jia Zhang and Chih-Chun Wang. On the rate-cost of gaussian linear control systems with random communication delays. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 2441–2445. IEEE, 2018.

[22] V. Kostina and Babak Hassibi. Rate-cost tradeoffs in control. In *Fifty-fourth Annual Allerton Conference*, pages 1157–1164, Allerton House IL, September 2016.

[23] T. Tanaka, K.H. Johansson, T. Oechtering, H. Sandberg, and M. Skoglund. Rate of prefix-free codes for LQG control systems. In *IEEE International Symposium on Information Theory*, pages 2399–2403, Barcelona, Spain, 2016.

[24] Photios A Stavrou, Jan Østergaard, Charalambos D Charalambous, and Milan Derpich. An upper bound to zero-delay rate distortion via kalman filtering for vector gaussian sources.

In *2017 IEEE Information Theory Workshop (ITW)*, pages 534–538. IEEE, 2017.

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

[26] Photios Stavrou and Jan Ostergaard. Fixed-rate zero-delay source coding for stationary vector-valued gauss-markov sources. In *Data Compression Conference, March 27-30 2018*, 2018.

[27] Photios Stavrou, Jan Ostergaard, and Charalambos Demetriou Charalambous. Zero-delay rate distortion via filtering for vector-valued gaussian sources. *IEEE Journal of Selected Topics in Signal Processing*, 2018.

[28] Alexey S Matveev and Andrey V Savkin. *Estimation and control over communication networks*. Springer Science & Business Media, 2009.

[29] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Dover Books on Electrical Engineering. Dover Publications, Mineola NY, 2012.

[30] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley and Sons Inc., New York, 2006.

[31] David Angeli. A Lyapunov approach to incremental stability properties. *IEEE Transactions on Automatic Control*, 47(3):410–421, 2002.

[32] Chun-Chia Huang and Robert R Bitmead. Escape time formulation of state estimation and stabilization with quantized intermittent communication. *Automatica*, 61:201–210, 2015.

[33] B.D.O. Anderson and J.B. Moore. Detectability and stabilizability of time-varying discrete-time linear systems. *SIAM J. Control Optimization*, 19(1):20–32, 1981.

[34] Salvatore Crisafulli. *Adaptive speech coding via feedback techniques*. PhD thesis, Australian National University, Canberra ACT Australia, 1992.

[35] D. Simon. *Optimal State Estimation: Kalman, H_∞, and nonlinear approaches*. John Wiley & Sons, Hoboken NJ, 2006.

[36] A. Doucet, J.F.G. de Freitas, and N.J. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, NY, 2001.

[37] J.F. Ferreira, J. Lobo, and J. Dias. Bayesian real-time perception algorithms on GPU. *Journal of Real-Time Image Processing*, 6(3):171–186, 2011.

[38] M. Loève. *Probability Theory vols 1 & 2*. Springer Verlag, Berlin, 1977.

[39] Panqanamala Ramana Kumar and Pravin Varaiya. *Stochastic systems: Estimation, identification, and adaptive control*, volume 75. SIAM, 2015.

[40] Frederick J Beutler. Dynamic programming: Deterministic and stochastic models (dimitri p. bertsekas). *SIAM Review*, 31(1):132, 1989.

[41] T. Mita. Optimal digital feeedback control systems counting computation time of control laws. *IEEE Transactions on Automatic Control*, 30(6):542–548, 1985.

[42] Vivek S Borkar and Sanjoy K Mitter. Lqg control with communication constraints. In *Communications, Computation, Control, and Signal Processing*, pages 365–373. Springer, 1997.

[43] Robert Gray. Vector quantization. *IEEE Assp Magazine*, 1(2):4–29, 1984.

[44] Chun-Chia Huang, Behrooz Amini, and Robert R Bitmead. Predictive coding and control. *IEEE Transactions on Control of Network Systems*, 6(2):906–918, 2018.

[45] Minyue Fu. Lack of separation principle for quantized linear quadratic gaussian control. *IEEE Transactions on Automatic Control*, 57(9):2385–2390, 2012.

[46] T Fischer. Optimal quantized control. *IEEE Transactions on Automatic Control*, 27(4):996–998, 1982.

[47] Girish N Nair and Robin J Evans. Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM Journal on Control and Optimization*, 43(2):413–436, 2004.

[48] Vijay Gupta, Amir F Dana, Richard M Murray, and Babak Hassibi. On the effect of quantization on performance at high rates. In *American Control Conference, 2006*, pages 6–pp. IEEE, 2006.

[49] Jennifer A Fulton, Robert R Bitmead, and Robert C Williamson. Sampling rate versus quantisation in speech coders. *Signal processing*, 56(3):209–218, 1997.

[50] Victoria Kostina and Babak Hassibi. Rate-cost tradeoffs in control. *IEEE Transactions on Automatic Control*, 64(11):4525–4540, 2019.

[51] Graham C Goodwin, Mauricio Esteban Cea Garrido, Arie Feuer, and David Q Mayne. On the use of one bit quantizers in networked control. *Automatica*, 50(4):1122–1127, 2014.

[52] Mauricio G Cea, GC Goodwin, Arie Feuer, and David Q Mayne. On the control rate versus quantizer-resolution trade off in networked control. *IFAC Proceedings Volumes*, 47(3):10343–10348, 2014.

[53] Peter E Caines. *Linear stochastic systems*, volume 77. SIAM, 2018.

[54] Torsten Söderström. *Discrete-time stochastic systems: estimation and control*. Springer Science & Business Media, 2012.