Linear Quadratic Gaussian Control with Quantized Feedback

Minyue Fu

Abstract—This paper studies a generalized linear quadratic Gaussian (LQG) control problem subject to the constraint that the feedback signal is quantized with a fixed bit rate. We show that state feedback control, state estimation and quantization can not be fully separated in general. Only a weak separation principle holds which converts the quantized LQG control problem into a quantized state estimation problem. We show that under high resolution quantization and the assumption that the output dimension does not exceed the input dimension, full separation of state feedback control, state estimation and quantization can be achieved.

I. INTRODUCTION

Quantized feedback control has received a lot of attention in recent years, due to the overwhelming need for network based control systems. This proposes many new challenges to the seemingly well-established linear control theory in questioning how to redesign control laws suitable for a networked environment. In this paper, we focus on the so-called quantized LQG control problem which is generalized from the standard LQG problem in discrete time but with the constraint that the feedback channel is a digital link with a fixed bit rate. We first look back at the history of the research on this problem and discuss many attempts to generalize the separation principle, some dated back to early 1960’s. We point out that many of these generalizations contain technical errors and/or misinterpretations. This leads us to a number of results on quantized LQG control, as briefed below:

1) A weak separation principle holds which states that optimal quantized LQG control can be achieved by separately designing state estimation, state feedback control and quantization. However, the separation is weak in two ways: i) The quantization criterion depends on the control cost function; ii) More seriously, optimal quantization can not be done by separately minimizing the quantization errors at different time instants. These weaknesses imply that optimal design for quantized LQG control is very complex numerically and is in huge contrast with the classical separation principle where state estimation is independent of the state feedback control and state estimation at each time instant can be done recursively without considering the future evolution of the system dynamics.

2) The consequence of the weak separation principle is that the quantized LQG problem becomes a quantized state estimation problem. In this problem, the output signal of a system needs to be quantized by a fixed rate quantizer and the quantized information is used to construct an estimate of a linear function of the state of the system, the desired control signal in our case, in a way to minimize a given distortion function. We point out that this can be viewed as a generalized vector quantization problem. We then use a linear predictive coding (LPC) type of approach to show that, under high resolution quantization and some mild rank condition, optimal quantization is done by using a memoryless quantizer. Using memoryless quantizers means that quantization can be done by considering each input sample separately. This result, together with the weak separation principle above, shows that a full separation principle holds for quantized LQG control under high resolution quantization and the mild rank condition. This rank condition essentially requires the dimension of output not to exceed the dimension of the input, which holds in particular for single-input-single-output systems.

II. PROBLEM FORMULATION

The quantized LQG problem we study is the same as the standard LQG control problem but with the constraint that the feedback signal must be quantized and transmitted over a digital link with a fixed bit rate, as depicted in Figure 1.

Fig. 1. Quantized LQG Control System

The system we consider is a discrete-time model given by

\[
\begin{align*}
x_{t+1} &= Ax_t + Bu_t + w_t \\
y_t &= Cx_t + v_t
\end{align*}
\]

(1)

where \(x_t \in \mathbb{R}^n\) is the state, \(u_t \in \mathbb{R}^m\) is the control input, \(y_t \in \mathbb{R}^p\) is the measured output, \(w_t \in \mathbb{R}^n\) and \(v_t \in \mathbb{R}^p\) are independent Gaussian random distributions with zero mean and covariances \(W_t > 0\) and \(V_t > 0\), respectively, and the initial state \(x_0\) is also assumed to be an independent zero-mean Gaussian distribution with covariance \(\Sigma_0\).

In the sequel, we denote \(z^t = \{z_0, z_1, \ldots, z_t\}\). The communication channel we consider in this paper is assumed to be a memoryless and error-free channel with a fixed transmission rate of \(R\) bits per sample. The output
signal \( y_t \) needs to be encoded first (as indicated by the ENC block in Fig 1) before transmission, and the received signal is decoded which is then used to construct a control signal \( u_t \) (as indicated by the DEC block in Fig. 1).

The encoder is required to be a causal mapping from the measured signal \( y_t \), i.e.,

\[
a_t = \alpha_t(y_t^t|a^{t-1})
\]

where \( \alpha(\cdot) \) takes values in a finite alphabet set \( A \) with size of \( 2^R \). Without loss of generality, we take \( A = \{1, 2, \ldots, 2^R\} \).

Similarly, the decoder is required to be a causal mapping from the received quantized signal, i.e.,

\[
u_t = \beta_t(\tilde{a}_t|a^{t-1})
\]

where \( \tilde{a}_t \) is the received version of \( a_t \). Because the channel is error free, \( \tilde{a}_t = a_t \), thus (3) can be rewritten as

\[
u_t = \beta_t(a_t|a^{t-1})
\]

We are interested in the following linear quadratic cost:

\[
J = \mathcal{E} \left[ x_t^T Q_T x_T + \sum_{t=0}^{T-1} x_t^T Q_t x_t + 2u_t^T H_t x_t + u_t^T S_t u_t \right]
\]

where \( \mathcal{E}[\cdot] \) is the expectation operator and \( Q_t = Q_t', S_t = S_t' \) and \( H_t \) are weighting matrices with

\[
S_t > 0, \quad Q_t - H_t S_t^{-1} H_t' \geq 0
\]

for all \( t = 0, 1, \ldots, T-1 \) and \( Q_T = Q_T' \geq 0 \).

The problem of quantized LQG control is to jointly design the quantizer and controller (or encoder and decoder) to minimize the cost \( J \), under the bit rate constraint.

### III. Literature Review

It is interesting to know that the quantized LQG control problem, with a virtually identical problem formulation as in Section II, has been actively studied for a long time. Lewis and Tou [6] (1965), Meier [7] (1965) and a monograph by Tou [8] (1963) were perhaps the earliest attempts on this problem. Larson [9] (1967) for the first time claimed that the well-known separation principle for LQG control can be generalized to quantized LQG control. But their claim is incorrect [9]. Fischer [11] (1982) revisited the quantized LQG control problem. He correctly pointed out that the optimal quantizer must be time-variant. By allowing time-variant quantization, Fisher claimed that the separation of control, estimation and quantization indeed holds. More specifically, he claimed that the optimal quantized LQG problem is solved by separately

1) performing the optimal state estimation (Kalman filtering) to produce a state estimate \( \hat{x}_t \);
2) generating the optimal control \( u_t = K_t \hat{x}_t \) to the standard LQG problem;
3) quantizing \( u_t \) to minimize a weighted quantization distortion \( \mathcal{E}[(u_t - \hat{u}_t^t)^2/\Omega_t(u_t - \hat{u}_t^t)] \) for some weighting matrix \( \Omega_t \) dependent on the cost function, where \( \hat{u}_t^t \) is the quantized version of \( u_t \).

Borkar and Mitter [1] (1997) approached the quantized LQG control problem under a slightly different setting (by assuming the full state being measured without noise, but allowing a certain type of transmission errors). It is claimed in [1] that if, instead of the output, an “innovation process” is encoded and transmitted, control, estimation and quantization can be separately designed to achieve optimal performance. It should be noted that the separation is conditioned on the specific choice of the quantization scheme (i.e., quantization of a particular “innovation process”). The paper did not discuss whether such a choice is optimal or not. Also, this method does not apply to unstable systems [2].

In the work of Tatikonda, Sahai and Mitter [2] (2004), the quantized LQG control problem is revisited once again also under the assumption of full state being measured. By using the work of Bar-Shalom and Tse [13] on the so-called neutrality (or no dual effect), a separation result is given in [2]. In the work of Matveev and Savkin [3] (2004), a restrictive decoder, called memoryless decoder for which (4) reduces to \( u_t = \beta_t(a_t) \), is used for quantized LQG control. It was shown in [3] that if using a memoryless decoder, separation of state feedback control, state estimation and quantization holds. We note that the use of memoryless decoders can degrade the control performance.

Despite all these claims, we will show via an example that the separation principle breaks down for quantized LQG control. That is, strictly speaking, the optimal control law can not be separated into independent state feedback control, state estimation and quantization problems. We point out that the lack of separation for quantized LQG has been recognized by, Bao et. al. [5] (2008), although no explicit examples are shown there.

### IV. Separation Principle and Lack of It

The core of the classical LQG control theory is the well-known separation principle which states that the optimal controller is given by \( u_t = K_t \hat{x}_t \), where \( K_t \) is the optimal control gain assuming that the true state is known, and \( \hat{x}_t \) is the optimal estimate of the state \( x_t \) based on \( y_t \). An important question to ask, in the presence of a bit rate constraint, is whether the separation principle generalizes or not.

We first recall the optimal control solution for \( K_t \) below:

\[
K_t = -(S_t + B'P_{t+1}B)^{-1}(B'P_{t+1}A + H_t)
\]

\[
P_t = Q_t + A'P_{t+1}A - K_t(S_t + B'P_{t+1}B)K_t'
\]

with \( P_T = Q_T \) and the optimal \( J \) for state feedback is [14]

\[
J_{LQ} = \text{tr}(P_0\Sigma_0) + \sum_{t=0}^{T-1} \text{tr}(W_tP_{t+1})
\]

It is also well known (see, e.g., [15]) that the optimal state estimate \( \hat{x}_t \) is obtained by the following Kalman filter:

\[
\hat{x}_{t+1|t} = A\hat{x}_t + Bu_t
\]

\[
\hat{x}_t = \hat{x}_{t|t-1} + E_tC'V_t^{-1}(y_t - C\hat{x}_{t|t-1})
\]

with \( \hat{x}_{0|t} = E[\hat{x}_0] = 0 \), where

\[
E_t = E[(x_t - \hat{x}_t)(x_t - \hat{x}_t)']
\]
is the state estimation error covariance given by:

\[ E_t = E_{t|t-1} - E_{t|t-1}C'(CE_{t|t-1}C' + V_t)^{-1}CE_{t|t-1} \]

\[ E_{t+1|t} = AE_tA' + ... \]

pre-computed and in the steady state they are constant. Thus, the next question is whether the quantization of \{z_t\}

\[ J_{LQG} = J_{LQ} + \sum_{t=0}^{T-1} \text{tr}(K_t'\Omega_tK_tE_t) \]

where

\[ \Omega_t = S_t + B'P_{t+1}B \]

**A. Weak Separation of Quantized LQG Control**

Returning to the quantized LQG control problem, we have the following basic result:

**Theorem 4.1:** Consider the quantized LQG control problem for the system (1), the cost function (5) and R-bit fixed-rate quantization. Denote

\[ z_t = K_t\hat{x}_t \]

and its quantized version by \( z_t^q \). Then, optimal quantized LQG control is achieved by choosing the encoder (sequence) \( \alpha_t \) to minimize the following distortion function:

\[ D = \sum_{t=0}^{T-1} \mathcal{E}[(z_t - z_t^q)'\Omega_t(z_t - z_t^q)] \]

The corresponding optimal controller and cost function are given by, respectively,

\[ u_t = z_t^q \]

\[ \min J = J_{LQG} + \min D \]

Moreover, given any encoder \( \{\alpha_t\} \) in (2), the optimal solution to \( u_t \) is given by

\[ u_t = \mathcal{E}[K_t\hat{x}_t|a^t] = K_t\mathcal{E}[\hat{x}_t|a^t] \]

**Proof:** The proof is essentially borrowed from [11]. Using the standard dynamic programming technique [15], the cost function \( J \) can be rewritten as

\[ J = J_{LQ} + \sum_{t=0}^{T-1} \mathcal{E}((K_t\hat{x}_t - u_t)'\Omega_t(K_t\hat{x}_t - u_t)) \]

Rewriting

\[ K_t\hat{x}_t = K_t\hat{x}_t + K_t(x_t - \hat{x}_t) \]

and using (10), we get

\[ J = J_{LQG} + D + 2\sum_{t=0}^{T-1} \mathcal{E}((K_t\hat{x}_t - u_t)'\Omega_t(x_t - \hat{x}_t)) \]

It is a well-known fact [15] that \( x_t - \hat{x}_t \) is uncorrelated to \( \hat{x}_t \) and \( y^t \). Since they are all Gaussian, \( x_t - \hat{x}_t \) is independent of \( \hat{x}_t \) and \( y^t \) and hence independent of \( u_t \) too because the latter depends only on \( y^t \). Therefore, the last term in \( J \) above is zero and we arrive at (17). The equation (18) is easily obtained by minimizing the term \( \mathcal{E}[(z_t - u_t)'\Omega_t(z_t - u_t)] \) in \( D \), conditioned on \( a^t \).

The next result shows that the quantized LQG problem is equivalent to a quantized state estimation problem. For this purpose, we consider the open-loop system of (1) as follows:

\[ \hat{x}_{t+1} = A\hat{x}_t + w_t \]

\[ \hat{y}_t = C\hat{x}_t + v_t \]

It is clear that \( x_t \) and \( \hat{x}_t \) are related by

\[ x_t = \hat{x}_t - \sum_{i=0}^{t-1} Bu_i \]

Consider the following distortion function

\[ \tilde{D} = \sum_{t=0}^{T-1} \mathcal{E}(\hat{x}_t - \tilde{x}_t^q)'\Omega_t(\hat{x}_t - \tilde{x}_t^q) \]

where \( \Pi_t \geq 0, \tilde{x}_t^q \) is the quantized \( \hat{x}_t \) and the associated encoder is given by

\[ a_t = \tilde{\alpha}_t(\tilde{y}_t|a^{t-1}) \]

which is a rate-\( R \) encoder. The quantized state estimation problem is to find an encoder (22) such that \( \tilde{D} \) is minimized.

**Theorem 4.2:** Consider the quantized LQG problem for the system (1) with the cost function (5) and fixed bit rate \( R \). Define

\[ \Pi_t = K_t'\Omega_tK_t \]

and let \( \{\tilde{\alpha}_t(\cdot)\} \) be the optimal encoder that minimizes the distortion function (21) for the associated quantized state estimation problem. Then, the optimal encoder (2) for the quantized LQG problem is given by

\[ \alpha_t(y_t|a^{t-1}) = \tilde{\alpha}_t(\tilde{y}_t|a^{t-1}) \]

with

\[ \tilde{y}_t = y_t - \sum_{i=0}^{t-1} CBu_i \]

Moreover, the minimum cost for the quantized LQG problem is given by

\[ \min J = J_{LQG} + \min \tilde{D} \]

and

\[ x_t^q = \tilde{x}_t^q + \sum_{i=0}^{t-1} Bu_i = \mathcal{E}[\tilde{x}_t|a^t] + \sum_{i=0}^{t-1} Bu_i \]

**Proof:** The proof follows from Theorems 4.1 and two simple facts below: 1) \( \tilde{y}_t \) is linear in \( a^{t-1} \); 2) \( u_t = K_tx_t^q \) with \( x_t^q \) fully determined by \( a^t \). These two facts collectively mean that \( y_t|a^t \) and \( \tilde{y}_t|a^t \) possess the same information. Hence, we can encode \( \tilde{y}_t|a^t \) instead of \( y_t|a^t \) and construct \( x_t^q \) from \( \tilde{x}_t^q \) as in (27) without affecting the distortion.

**B. Lack of Separation for Quantization**

From Theorems 4.1-4.2, we understand that the quantized LQG control problem boils down to quantizing the sequence \{z_t\} or \{\hat{x}_t\}. We see that the distortion function \( D \) depends on the cost function \( J \), but the relevant parameters \( K_t \) and \( \Omega_t \) can be pre-computed and in the steady state they are constant. Thus, the next question is whether the quantization of \{z_t\}
can be further separately so that each $z_t$ can be separately quantized. This question is made more precise below. Denoting

$$D_t = E[(z_t - u_t)^r \Omega_t (z_t - u_t)]$$

(28)

then the distortion function can be split into two terms:

$$D = \sum_{t=0}^{T-1} D_t$$

(29)

The function $\alpha_t$ needs to be designed to minimize the second term above called the distortion-to-go. The specific separation question we ask is whether the optimal encoder sequence $\{\alpha_t\}$ for minimizing $D$ can be obtained by taking each $\alpha_t$ separately to minimize the distortion $D_t$ only.

Unfortunately, this type of separation is not possible, as demonstrated by the example below.

**Example 4.1:** The example we consider is a scalar system without any process or measurement noise:

$$x_{t+1} = x_t + u_t, \quad x_0 \sim N(0, 1)$$
$$y_t = x_t$$

(30)

where $N(0, 1)$ is the normalized Gaussian distribution (with zero mean and unity variance). The cost function is given by

$$J = E[Q_1 x_0^2 + Q_2 x_2^2 + Q_3 x_2^2 + R_0 u_0^2 + R_1 u_1^2 + R_2 u_2^2]$$

(31)

A single-bit quantizer is to be used.

Using (14), (15) and (17), we rewrite $J = J_{LQG} + D$ with

$$D = E[\Pi_0(x_0 - x_0^q)^2 + \Pi_1(x_1 - x_1^q)^2 + \Pi_2(x_2 - x_2^q)^2]$$

Note that $\{\Pi_t\}$ is a function of $\{Q_t\}$ and $\{R_t\}$. Conversely, we can choose $\{Q_t\}$ and $\{R_t\}$ to make any positive $\{\Pi_t\}$ we want. In particular, we will take $\Pi_0 = \Pi_1 = 1$ and leave $\Pi_2 > 0$ as a free parameter. Defining

$$\rho_0 = x_0^q$$
$$\rho_1 = x_1^q - K_0 x_0^q$$
$$\rho_2 = x_2^q - K_1 x_1^q - K_0 x_0^q$$

and using $x_{t+1} = x_t + K_t x_t^q$, $D$ can be rewritten as

$$D = E[(x_0 - \rho_0)^2 + (x_1 - \rho_1)^2 + \Pi_2(x_2 - \rho_2)^2]$$

(32)

Therefore, for this particular example, the quantized LQG problem becomes a quantization problem for $x_0 \sim N(0, 1)$ with distortion in (32). We can view $\{\rho_t\}$ as a sequence of successive quantized estimates of $x_0$, i.e., $\rho_t = E[x_0^q|a^t]$.

Figure 2 shows how the quantization works: At $t = 0$, the range of $x_0$ is split into two quantization intervals: $(-\infty, 0)$ and $(0, \infty)$. At $t = 1$, each of the above intervals is divided into two intervals: The interval $(0, \infty)$ is divided into $(0, i_2)$ and $(i_2, \infty)$, and $(-\infty, 0)$ divided into $(-\infty, -i_2)$ and $(-i_2, 0)$. At $t = 2$, each of the above intervals is further divided into two. Thus, we will have 4 intervals on the positive side: $(0, i_1), (i_1, i_2), (i_2, i_3), (i_3, \infty)$, and the negative side is mirror imaged. It can be shown that this symmetric structure is optimal for any $S_2$.

At $t = 0$:

$$x_{0+1} = x_0 + u_0, \quad x_0 \sim N(0, 1)$$
$$y_0 = x_0$$

At $t = 1$:

$$x_{1+1} = x_1 + K_1 x_1^q, \quad x_1 = x_0 + K_0 x_0^q$$
$$y_1 = x_1$$

At $t = 2$:

$$x_{2+1} = x_2 + K_2 x_2^q, \quad x_2 = x_1 + K_1 x_1^q - K_0 x_0^q$$
$$y_2 = x_2$$

Fig. 2. Example of Quantized LQG

The separation question we asked before becomes the following: Suppose, at time $t$, $y_t$ is encoded to minimize $E[(x_0 - \rho_t^q)^2|a^t]$ at each time $t$. Will such an encoder lead to the optimal $D$? The answer turns out to be negative. If $\Pi_2 \to 0$, to minimize $D$, the optimal value for $i_2 = 0.9816$. This value is also optimal for minimizing $D_1$ (the second term in (32)). The corresponding optimal values for $i_1$ and $i_3$ are 0.4709 and 1.6942, respectively. If $\Pi_2 \to \infty$, to minimize $D$, the optimal value for $i_2 \to 1.05$, and the corresponding optimal values for $i_1$ and $i_3$ are 0.5006 and 1.7470, respectively. If $\Pi_2$ is very large and we use the optimal value of $i_2$ for minimizing $D_1$, there will be some increase in the distortion (around 0.8%).

V. QUANTIZED STATE ESTIMATION

We now study the quantized state estimation problem. The system we consider is given by

$$x_{t+1} = A x_t + u_t$$
$$y_t = C x_t + v_t$$

(33)

with $x_0, \{w_t\}, \{v_t\}$ being independent Gaussian random variables as before. Let $\hat{x}_t$ be the optimal (Kalman) estimate of $x_t$ and consider $z_t = K_t \hat{x}_t$ for some given $K_t$. The task of quantized state estimation is to encode $\{y_t\}$ (or $\{z_t\}$ indirectly) using (4) with fixed bit rate $R$ to minimize the following distortion function

$$D = \sum_{t=0}^{T-1} E[(z_t - z_t^q)^2 | \Omega_t (z_t - z_t^q)]$$

(34)

for some given $\Omega_t$, where $z_t^q$ is the quantized $z_t$.

A. Special Case: “White” Signal

We first study a special case where $\{z_t\}$ is an independent sequence of random variables, but different distributions are allowed at different $t$. We may consider this sequence to be generated by (33) with $A = 0$. In this case, $z_t = K_t \hat{x}_t$ with

$$\hat{x}_t = \Sigma_t C' (C \Sigma_t C' + V_t)^{-1} y_t$$

(35)
We have the following simple result which shows that a memoryless quantizer is optimal, i.e., each $z_t$ can be quantized independently. The result is somewhat surprising because in the standard quantization setting where there is no causality constraint, it is well known that memoryless quantization is only suboptimal (hence the need for vector quantization).

**Theorem 5.1:** Suppose $\{z_t\}$ is an independent sequence of random variables with probability density functions $\{f_t\}$. Then, the optimal quantizer that minimizes $D$ has the following memoryless structure:

$$a_t = \alpha_t(z_t); \quad z_t^q = \mathcal{E}[z_t|a_t]$$

with $\alpha_t$ chosen to minimize

$$D_t = \mathcal{E}[(z_t - z_t^q)^2]$$

**Proof:** The independence of $\{z_t\}$ implies that the a posteriori probability density function of $z_t$ is still $f_t$, not altered upon receiving $a_t$. For the same reason, quantization of $z_t$ has no effect on the future distortion terms $D_{t+\tau}, \tau > t$. Hence, the optimal quantizer must be the one that minimizes $D_t$ only, thus it is a memoryless quantizer.

\[\Box\]

**B. General Case: Colored Signal**

We return to the general case where $\{z_t\}$ is not an independent sequence. Motivated by Theorem 5.1, we may be tempted to consider a “whitening” approach, as depicted in Figure 3. In this approach, $z_t$ is “whitened” first before quantization. This can be done by passing $\{z_t\}$ through a whitening filter $F$ to generate a white sequence $\{u_t\}$ for quantization (Scheme (a)) or, alternatively, by quantizing the innovation signal

$$e_t = y_t - C\hat{x}_{t|t-1}$$

from the Kalman filter (9) directly (Scheme (b)). The quantized signal is then used to “reconstruct” the intended signal by ignoring the quantization error.

**Example 5.1:** Let

$$z_0 = n_0; \quad z_1 = n_1 + n_0$$

with $n_0$ and $n_1$ independent and uniformly distributed in $[0, 1]$. Let

$$D = \mathcal{E}[(z_0 - z_0^q)^2 + (z_1 - z_1^q)^2]$$

An $L$-level quantizer is to be used for a large $L$.

It is easy to “whiten” $z_t$ to get

$$n_0 = z_0; \quad n_1 = z_1 - z_0$$

If we quantize $\{n_t\}$, then, following Theorem 5.1, $n_t$ can be quantized independently. Since $u_t$ is uniformly distributed, a uniform quantizer is known to be optimal and the distortion is given by [17]

$$\mathcal{E}[(n_t - n_t^q)^2] = \delta^2/12$$

with $\delta = 1/L$. The corresponding distortion $D$ equals

$$D = \mathcal{E}[(n_0 - n_0^q)^2 + (n_1 + n_0 - n_1^q - n_0^q)^2] = \mathcal{E}[2(n_0 - n_0^q)^2 + (n_1 - n_1^q)^2] = \delta^2/4$$

Now consider the alternative quantization scheme where $z_0 = n_0$ is quantized using a uniform quantizer as before which gives $D_0 = \delta^2/12$, but $z_1$ is quantized differently as follows. At time $t = 1$, $z_0^q = n_0^q$ is known to lie uniformly in $[n_0^q - \delta/2, n_0^q + \delta/2]$. Then, $z_1 = n_1 + n_0$ has probability density function shown in Figure 4 (without the offset of $n_0^q$). Let $[n_0^q - \delta/2, n_0^q + \delta/2]$ and $[1 + n_0^q - \delta/2, 1 + n_0^q + \delta/2]$ to be two quantization intervals and take the remaining $L-2$ intervals to be uniform in $[n_0^q + \delta/2, 1 - n_0^q - \delta/2]$. Then, the distortion for $z_1$ is computed to be

$$D_1 = \mathcal{E}[(z_1 - z_1^q)^2] = \frac{(1 - \delta)^2}{12(L - 2)^2} (1 - \delta) + \frac{\delta^2}{18}$$

Combining it with the distortion for $z_0$, we get

$$D = \delta^2/6 + O(\delta^3)$$

where $O(\delta^3)$ involves only 3rd order terms of $\delta$. It is clear that when $L$ is large, this distortion is smaller than (43).

**Remark 5.1:** We should not confuse the whitening approach with the LPC approach. The latter quantizes the prediction error

$$e_t = z_t - \hat{z}_{t|t-1}$$

where $\hat{z}_{t|t-1}$ is the prediction of $z_t$ given by

$$\hat{z}_{t|t-1} = \mathcal{E}[z_t|e_t^q, 0 < \tau < t]$$

Because $e_t$ contains the same information as $z_t$ at time $t$, quantizing $e_t$ is equivalent to quantizing $z_t$. The key to the
LPC approach is that the previous quantization values are used in constructing the input to the quantizer, which is not the case in the whitening approach. The alternative approach in the example uses LPC, which explains why it is better than the whitening approach.

C. High Resolution Quantization: Separation Principle

We continue with the general case where \( \{z_t\} \) is a colored sequence and we will show that, under the assumption of high resolution quantization (i.e., \( R \) being large) and a mild rank condition, the optimal quantizer has a very simple separable structure. This is obtained using the LPC approach.

We first rewrite the Kalman filter (9) (with \( u_t = 0 \)) as

\[
\hat{x}_t = \hat{x}_{t|t-1}^q + \Gamma_t e_t + (\hat{x}_{t|t-1}^q - \hat{x}_{t|t-1}^q)
\]

(48)

where \( \Gamma_t = E_t C'V_t^{-1} \) and \( e_t \) is given in (38). Instead of quantizing \( z_t \) directly, we consider quantizing \( e_t \) using a memoryless quantizer and take

\[
\hat{x}_t^q = \hat{x}_{t|t-1}^q + e_t^q, \quad z_t^q = K_t \hat{x}_t^q
\]

(50)

Since \( e_t \) and \( \hat{x}_{t|t-1}^q - \hat{x}_{t|t-1}^q \) are independent, quantizing \( e_t \) will always yield a larger distortion for \( \hat{x}_t \) (hence for \( z_t \)) than in the case when \( \hat{x}_{t|t-1}^q - \hat{x}_{t|t-1}^q = 0 \).

We first mention a very important result on vector quantization by Zador [19] (1963) (also see [17], [18]) which states that the minimum distortion \( \mathcal{E}[\|x - x^q\|^2] \) for any \( k \)-dimensional random source \( x \) by a fixed-rate quantizer with bit rate \( R \) has the form

\[
D(R) \approx b_k \|f_x\|_k/(k+2) 2^{-2R}
\]

(51)

where \( f_x \) is the probability density function of \( x \),

\[
\|f_x\|_k/(k+2) = \left( \int_{\mathbb{R}^k} f_x^{k/(k+2)}(x) dx \right)^{(k+2)/k}
\]

(52)

and \( b_k \) is a term independent of the source \( x \), representing how well cells can be packed in \( k \)-dimensional space for the given distortion measure. The approximation \( \approx \) in (51) means that the approximation error is in an order smaller than \( 2^{-2R} \) as \( R \to \infty \). But the approximation is known to be very accurate even for relatively small values of \( R \) [20].

Using Zador’s formula (51), we obtain the main result of this section which, together with Theorems 4.1-4.2, establishes a complete separation for quantized LQG control under high resolution quantization and a mild rank condition. The proof is omitted due to page limit.

Theorem 5.2: Consider the quantized state estimation problem associated with the system (33) and the distortion function \( D \) in (34). Suppose \( K_t\Gamma_t \) has full column rank, where \( \Gamma_t = E_t C'V_t^{-1} \) as previously defined. Then, high resolution quantization of \( z_t \) is optimally done by quantizing \( e_t \) in (49) using a memoryless quantizer for the following distortion function:

\[
D_t^e = \mathcal{E}[(e_t - e_t^q)^tK_t^t\Omega_tK_t(e_t - e_t^q)]
\]

(53)

The minimum distortion \( D_t^e \) is approximately the same as the minimum distortion by quantizing \( e_t \) using a memoryless quantizer with the following distortion function:

\[
D_t^e = \mathcal{E}[(e_t - e_t^q)^tK_t^t\Omega_tK_t(e_t - e_t^q)]
\]

Hence, the minimum distortion \( D \) is approximated by

\[
\min D \approx \min \sum_{t=0}^{T-1} D_t^e
\]

(55)

with the approximation error in the order smaller than \( 2^{-2R} \).

REFERENCES