Robust Predictive Quantization: Analysis and Design Via Convex Optimization

Alyson K. Fletcher, *Member, IEEE*, Sundeep Rangan, Vivek K Goyal, *Senior Member, IEEE*, and Kannan Ramchandran, *Fellow, IEEE*

***Abstract*—Predictive quantization is a simple and effective method for encoding slowly-varying signals that is widely used in speech and audio coding. It has been known qualitatively that leaving correlation in the encoded samples can lead to improved estimation at the decoder when encoded samples are subject to erasure. However, performance estimation in this case has required Monte Carlo simulation. Provided here is a novel method for efficiently computing the mean-squared error performance of a predictive quantization system with erasures via a convex optimization with linear matrix inequality constraints. The method is based on jump linear system modeling and applies to any autoregressive moving average (ARMA) signal source and any erasure channel described by an aperiodic and irreducible Markov chain. In addition to this quantification for a given encoder filter, a method is presented to design the encoder filter to minimize the reconstruction error. Optimization of the encoder filter is a nonconvex problem, but we are able to parameterize with a single scalar a set of encoder filters that yield low MSE. The design method reduces the prediction gain in the filter, leaving the redundancy in the signal for robustness. This illuminates the basic tradeoff between compression and robustness.**

***Index Terms*—Differential pulse code modulation, erasure channels, joint source-channel coding, linear matrix inequalities.**

I. Introduction

PREDICTIVE quantization is one of the most widely-used methods for encoding time-varying signals. It essentially quantizes changes in the signal from one sample to the next, rather than quantizing the samples themselves. If the signal is slowly varying relative to the sample rate, predictive quantization can result in a significant reduction in distortion for a fixed number of bits per sample—a large *coding gain* [1]. Due to its effectiveness and simplicity, predictive quantization is the basis of most speech encoders [2]. It is also used in some audio coders [3] and, in a sense, in motion-compensated video coding [4].

Manuscript received February 1, 2007; revised September 1, 2007. This work was supported in part by a University of California President's Postdoctoral Fellowship and the Texas Instruments Leadership University Program. This work was presented in part at the IEEE International Symposium on Information Theory, Chicago, IL, June/July 2004 and the IEEE International Conference on Image Processing, Singapore, October 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhi-Quan (Tom) Luo.

A. K. Fletcher and K. Ramchandran are with the University of California, Berkeley, CA 94720 USA (e-mail: alyson@eecs.berkeley.edu; kannanr@eecs.berkeley.edu).

S. Rangan is with Qualcomm Flarion Technologies, Bedminster, NJ 07921 USA (e-mail: s.rangan@qualcomm.com).

V. K. Goyal is with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: vgoyal@mit.edu).

Digital Object Identifier 10.1109/JSTSP.2007.910622

A prohibitive weakness of predictive quantization is its relative lack of robustness to losses (or erasures) of quantized samples. Since predictive quantization essentially encodes only changes in the signal, a loss of any one quantized sample results in errors that propagate into future samples.

For an ergodic channel with known expected probability of erasure, an information-theoretically optimal approach is to use a block erasure-correcting code to mitigate the losses [5]. If the probability of erasure is unknown or the erasures are adversarial, a rateless code could be effective for such an application [6]. However, both of these approaches induce added end-to-end delay because coding is done over blocks. Predictive quantization is attractive because it adds no delay.

The purpose of this paper is to consider predictive quantization with losses, where the quantized samples must be transmitted without channel coding and the decoder must reconstruct the signal without additional delay. We develop a novel approach to quantifying the performance of these systems that uses jump linear state space systems and convex optimization with linear matrix inequalities. We also present a new design method that optimizes the encoder to minimize the distortion given the loss statistics. Such performance computation and encoder optimization previously required system simulation.

A. Jump Linear Systems

Jump linear systems are linear state-space systems with time-varying dynamics governed by a finite-state Markov chain. Such systems have been studied as early as the 1960s, but gained significant interest in the 1990s, when it was shown in [7] that they can be analyzed via a convex optimization approach known as linear matrix inequalities (LMIs). LMI-constrained optimizations and their application to jump linear systems are summarized in several books such as [8] and [9].

To describe the lossy predictive quantization problem with a jump linear system, this paper models the source signal to be quantized and the encoder prediction filter with linear state-space systems. The quantization noise is approximated as additive white and Gaussian. The channel between the encoder and decoder is then modeled as a random erasure channel where the losses of the quantized samples are described by a finite-state Markov chain. The resulting system—signal generation, encoding, and channel—can be represented with a single jump linear state-space model.

B. Preview of Main Results

Using jump linear estimation results in [10]–[12], we show that the minimum average reconstruction error achievable at the decoder can be computed with a simple LMI-constrained optimization. A jump linear reconstruction filter at the decoder that achieves the minimal reconstruction error is also presented.

The proposed LMI method is quite general in that it applies to autoregressive moving average (ARMA) signal models of any order, encoder filters of any order, and any Markov erasure channel. The Markov modeling is particularly important: With channel coding and sufficient interleaving, the overall system performance is governed only by the average loss rate. However, with uncoded transmission, it is necessary to model the exact channel dynamics and time correlations between losses, and Markov models can capture such dynamics easily.

Using the LMI method to evaluate the effect of losses, we propose a method for optimizing the encoder filter to minimize the distortion given the loss statistics. The method essentially attempts to reduce the prediction gain in the filter, thereby leaving redundancy in the quantized samples that can be exploited at the decoder to overcome the effect of losses.

The approach illuminates a general tradeoff between compression and robustness. Compression inherently removes redundancy, reducing the required data rate, but also making the signal more vulnerable to losses. The results in this paper can be seen as a method for optimizing this tradeoff for linear predictive quantization by controlling the prediction gain in the encoder.

A final, and more minor, contribution of the paper is that a number of standard results for lossless predictive quantizer design are rederived in state-space form; these state-space formulas seem to not be widely known. The state-space method can incorporate the effect of closed-loop quantization noise more easily than frequency-domain methods such as [13] and [14]. Also, the state-space results show an interesting connection to lossy estimation with no quantization as studied in [15].

C. Previous Work

In making predictive quantization robust to losses of transmitted data, the prevailing methods are focused entirely on the decoder. When the source is a stationary Gaussian process, the optimal estimator is clearly a Kalman filter. Authors such as Chen and Chen [16] and Gündüzhan and Momtahan [17] have extended this to speech coding and have obtained significant improvements over simpler interpolation techniques. While motion compensation in video coding is more complicated than linear prediction, the extensive literature on error concealment in video coding is also related. See [18], [19] for surveys and [20] for recent techniques employing Kalman filtering. Another line of related research uses residual source redundancy to aid in channel decoding [21]–[25]. These works are all focused on discrete-valued sources sent over discrete memoryless channels, and most of them assume retransmission of incorrectly decoded blocks. Here we consider the end-to-end communication problem, with a continuous-valued source and potentially complicated source and channel dynamics.

D. Comments on the Scope of the Results

The results presented here allow an ARMA signal model and Markov erasure channel, each of arbitrary order. Our convex optimization framework can be applied with greater generality. In particular, the source signal can have any finite number of modes, each described by an ARMA model, as long as the modes form a Markov chain and the transitions in this Markov chain are provided as side information. Also, the channel can be generalized to have any linear, additive white noise description in each Markov chain state. We have omitted these more general formulations for clarity. Amongst the more general formulations in [11] is an optimization of filters for multiple description coding of speech as proposed in [26].

The main limitation of this work is the need for some simplifying assumptions regarding the distribution of quantization noise (see Section III-B). These are merely approximations to the true characteristics of quantization noise that simplify our developments. We believe that finer characterization of quantization noise would have significant effect only if the estimators were allowed to be nonlinear. To justify the model more rigorously, we could employ vector quantizers and results from [27]. This would unnecessarily complicate our development.

E. Organization of the Paper

Section II defines jump linear systems and presents the key theoretical result that relates minimum estimation error in a jump linear system to an LMI-constrained convex optimization. Section III then explains how predictive quantization can be modeled in state space. As shown in Section IV, transmission of quantized prediction errors over a Markov erasure channel turns the overall system into a jump linear system for which the result of Section II applies to determine the encoding performance. The effect of optimizing the encoding is shown through numerical examples in Section V.

II. LMI Solutions to Optimal Jump Linear Estimation

A *jump linear* system is a linear state-space system with time variations driven by a Markov chain. A complete discussion of such systems can be found in [9], [11]. Here, we present an LMI-constrained optimization solution to the specific jump linear estimation problem arising in predictive quantizer design. Our result is closely related to the dual control problem presented in [10]. A complete discussion and proofs of the broader estimation result this is based on can be found in [11].

A discrete-time jump linear system is a state-space system of the form

$$\begin{aligned} x[k+1] &= A_{\theta[k]}x[k] + B_{\theta[k]}w[k], \\ z[k] &= C_{1,\theta[k]}x[k] + D_{1,\theta[k]}w[k], \\ y[k] &= C_{2,\theta[k]}x[k] + D_{2,\theta[k]}w[k] \end{aligned} \quad (1)$$

where $w[k]$ is unit-variance, zero-mean white noise, $x[k]$ is an internal state, $z[k]$ is an output signal to be estimated, and $y[k]$ is an observed signal. The parameter $\theta[k]$ is a discrete-time Markov chain with a finite number of possible states: $\theta[k] \in \{0,\ldots,M-1\}$. Thus, a jump linear system is a standard linear state-space system, where each system matrix can, at any time, take on one of M possible values. The term "jump" indicates that changes in the Markov state $\theta[k]$ result in discrete changes, or jumps, in the system dynamics. We will denote the Markov chain transition probabilities by $p_{ij} = \mathbf{Pr}(\theta[k+1]=j \mid \theta[k]=i)$ and assume that the Markov chain is aperiodic and irreducible. Thus, there is a unique stationary distribution $q_i = \mathbf{Pr}(\theta[k] = i)$ satisfying the equations $q_j = \sum_{i=0}^{M-1} p_{ij} q_i$ for $j = 0, \ldots, M-1$.

We consider the estimation of the unknown signal $z[k]$ from the observed signal $y[k]$. We assume that the estimator knows the Markov state $\theta[k]$; the estimation problem for the case when $\theta[k]$ is unknown is a more difficult, nonlinear estimation problem considered, e.g., in [28]–[30].

Under the assumption that $\theta[k]$ is known, the estimator essentially sees a linear system with known time variations. Consequently, the optimal estimate for $z[k]$ can be computed with a standard Kalman filter [31]. (To establish notation, standard Kalman filtering is reviewed in Appendix A.) However, without actual simulation of the filter, there is no simple method to compute the average performance of the estimator as a function of the Markov statistics. Also, the estimator requires a Riccati equation update with each time sample that may be computationally difficult.

We thus consider a suboptimal jump-linear estimator of the form: when $\theta[k] = i$,

$$\begin{aligned} \hat{x}[k+1] &= A_i \hat{x}[k] + L_{i1}(y[k] - C_{i2}\hat{x}[k]), \\ \hat{z}[k] &= C_{i1}\hat{x}[k] + L_{i2}(y[k] - C_{i2}\hat{x}[k]). \end{aligned} \tag{2}$$

The estimator (2) is itself a jump linear system whose input is the observed signal $y[k]$ and output is the estimate $\hat{z}[k]$. The system is defined by the two sets of matrices: $L_1 = (L_{0,1}, \ldots, L_{M-1,1})$ and $L_2 = (L_{0,2}, \ldots, L_{M-1,2})$. The estimator is similar in form to the causal Kalman estimator [Appendix A, (39)], except that the gain matrices are time varying via the Markov state $\theta[k]$.

Given an estimator of the form (2) for the system (1), we can define the error variance

$$\sigma^2(L_1, L_2) = \mathbf{E}\|z[k] - \hat{z}[k]\|^2$$

where the dependence on L_1 and L_2 is through the estimate $\hat{z}[k]$ in (2). The goal is to find gain matrices L_1 and L_2 to minimize $\sigma^2(L_1, L_2)$:

$$\min_{L_1, L_2} \sigma^2(L_1, L_2). \tag{3}$$

The following result from [11], [12] shows that the optimization can be solved with an LMI-constrained optimization. The paper [12] also precisely defines *MS stabilizing*, a condition on the gain matrices L_1 and L_2 to guarantee that the state estimates are bounded.

Theorem 1: Consider the jump linear estimation problem above.

a) Suppose that $[C_{i1}' \; C_{i2}']$ is onto for all i, and suppose that there exist matrices W_i and V_i, $i = 0, \ldots, M-1$, partitioned as

$$W_i = \begin{bmatrix} W_{i1} & W_{i2} \\ W_{i2}' & W_{i3} \end{bmatrix}, \quad V_i = \begin{bmatrix} I & V_{i2} \\ V_{i2}' & V_{i3} \end{bmatrix} \tag{4}$$

satisfying

$$\begin{aligned} W_{i1} &\geq [A_i' \; C_{i2}'] \bar{W}_i [A_i' \; C_{i2}']' + [C_{i1}' \; C_{i2}'] V_i [C_{i1}' \; C_{i2}']', \\ \bar{W}_i &\geq 0, \\ V_i &\geq 0 \end{aligned} \tag{5}$$

where $\bar{W}_i$ is defined by

$$\bar{W}_i = \begin{bmatrix} \bar{W}_{i1} & W_{i2} \\ W_{i2}' & W_{i3} \end{bmatrix} \tag{6}$$

and

$$\bar{W}_{i1} = \sum_{j=0}^{M-1} p_{ij} W_{j1}. \tag{7}$$

Then, $\bar{W}_{i1} > 0$ for all i. Also, if we define

$$L_{i1} = -\bar{W}_{i1}^{-1} W_{i2}, \quad L_{i2} = -V_{i2} \tag{8}$$

the set of matrices (L_1, L_2) is MS stabilizing and the mean-squared error is bounded by

$$\sigma^2(L_1, L_2) \leq \sum_{i=0}^{M-1} q_i \mathbf{Tr}(E_i' \bar{W}_i E_i + F_i' V_i F_i), \tag{9}$$

where $E_i = [B_i' \; D_{i2}']'$ and $F_i = [D_{i1}' \; D_{i2}']'$.

b) Conversely, for any set of MS stabilizing gain matrices (L_1, L_2), there must exist matrices W_i and V_i satisfying (5) and

$$\sum_{i=0}^{M-1} q_i \mathbf{Tr}(E_i' \bar{W}_i E_i + F_i' V_i F_i) \leq \sigma^2(L_1, L_2). \tag{10}$$

Combining parts (a) and (b) of Theorem 1, we see that the minimum estimation error is given by

$$\min_{L_1, L_2} \sigma^2(L_1, L_2) = \min_{W_i, V_i} \sum_{i=0}^{M-1} q_i \mathbf{Tr}(E_i' \bar{W}_i E_i + F_i' V_i F_i) \tag{11}$$

where the first minimization is over MS stabilizing gain matrices (L_1, L_2), and the second minimization is over matrices W_i and V_i satisfying (4) and (5). For a fixed set of transition probabilities, p_{ij}, the objective function (11) and constraint (5) are linear in the variables W_i and V_i. Consequently, the optimization can be solved with LMI constraints, thus providing a simple way to optimize the jump linear estimator.

III. State-Space Modeling of Predictive Quantization

A. System Model

Since predictive quantization is traditionally analyzed in the frequency domain, we will need to first rederive some standard results in state space. For simplicity, this section considers only predictive quantization without losses. The lossy case, which is our main interest, will be considered in Section IV.

Fig. 1 shows the model we will use for the predictive quantizer encoder and decoder. The signal to be quantized is called the "source" and denoted by $z[k]$. The predictive quantizer encoder consists of a linear filter H_{enc}, in feedback with a scalar quantizer $Q(\,\cdot\,)$. The filter H_{enc} will be called the *encoder filter*

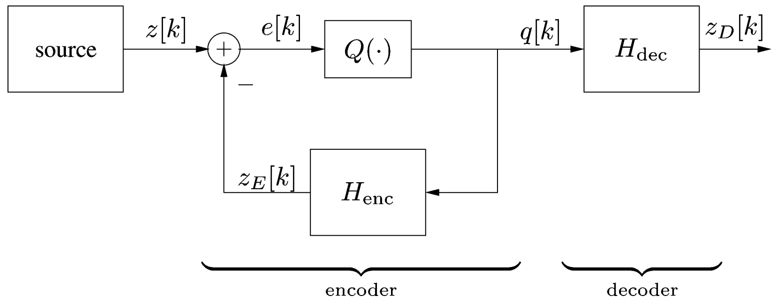

Fig. 1. Predictive quantizer encoder and decoder with a general higher-order filter.

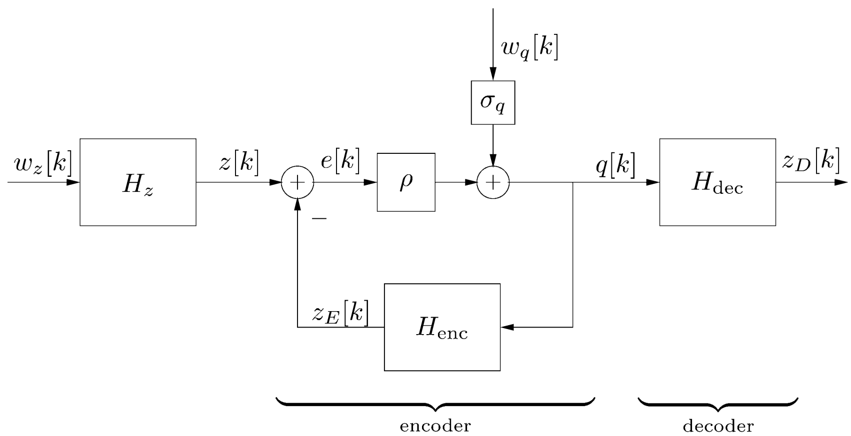

Fig. 2. Linear model for the predictive quantizer system. The quantizer is replaced by a linear gain and additive noise, and the source signal, $z[k]$, is described as filtered white noise.

and its output is denoted $z_E[k]$. The quantizer output sample sequence is denoted $q[k]$. The decoder is represented by the linear filter H_{dec}, which operates on the quantizer samples $q[k]$ to produce the final decoder output is denoted $z_D[k]$. The decoder output $z_D[k]$ is the final estimate of the original signal $z[k]$, so ideally $z_D[k]$ is close to $z[k]$.

B. Linear Quantizer Model

Due to the nonlinear nature of the scalar quantizer, an exact analysis of the predictive quantizer system is difficult. The classical approach to deal with the nonlinearity is to approximate the quantizer as a linear gain with additive white noise (AWN). One such model is provided by the following lemma.

Lemma 1: Consider the predictive quantizer system in Fig. 1. Suppose that the closed-loop system is well-posed and stable and all the resulting signals are wide-sense stationary random processes with zero mean. In addition, suppose that the scalar quantizer function $Q(\,\cdot\,)$ is designed such that

$$\mathbf{E}(e[k]\,|\,Q(e[k])) = Q(e[k]) = q[k]. \tag{12}$$

Also, let

$$\beta = \frac{\mathbf{E}|e[k]-q[k]|^2}{\mathbf{E}|e[k]|^2}. \tag{13}$$

It follows that the quantizer output can then be written (as shown in Fig. 2)

$$q[k] = \rho e[k] + \sigma_q w_q[k] \tag{14}$$

where $w_q[k]$ is a zero-mean, unit-variance signal, uncorrelated with $e[k]$, and

$$\rho = 1-\beta,$$
$$\sigma_q^2 = \beta(1-\beta)\mathbf{E}|e[k]|^2. \tag{15}$$

Proof: See Appendix C-1.

The lemma shows that the quantizer can be described as a linear gain ρ, along with additive noise $w_q[k]$ uncorrelated with the quantizer input $e[k]$. Observe that since $w_q[k]$ is normalized to have unit variance, the variance of the additive noise is σ_q^2.

The assumption (12) is that, for each partition region, the function Q will output the conditional mean. This condition is satisfied by optimal quantizers, or more generally, whenever the quantization decoder mapping is optimal, regardless of whether the encoder mapping is optimal [1]. Moreover, the condition will hold approximately for uniform quantizers at high rates.

The factor β in (13) is a proportionality constant between the quantizer input variance and quantizer error variance. The parameter can thus be seen as a measure of the quantizer's relative accuracy. In [1], the factor $1/\beta$ is called the *coding gain* of the quantizer. Following this terminology, we will call β either the *inverse coding gain* or *coding loss*.

In general, β will depend on the quantizer input distribution, specific scalar quantizer design and number of bits per sample. In particular, it is independent of the encoder and decoder filters, and can therefore be treated as a design constant. For example, suppose the input to the quantizer is well-approximated as a Gaussian scalar random variable and the quantizer $Q(\cdot)$ is a uniform scalar quantizer. Then, a well-known analysis by Bucklew and Gallagher [32] shows that $\beta \approx KR2^{-2R}$, where R is the quantizer rate (i.e., number of bits per sample), $K > 0$ is a constant and the relative error in the approximation becomes negligible for high rates R. If the quantized samples are then entropy coded, it is shown by Gish and Pierce [33] that the coding loss β is reduced to $\beta \approx (\pi e/6)2^{-2R}$.

While Lemma 1 provides a simple additive noise model for the quantizer, we require two further approximations for the linear modeling.

a) At each time sample k, the quantizer noise $w_q[k]$ constructed in Lemma 1 is guaranteed to be uncorrelated with the quantizer input $e[k]$. However, in the linear modeling we need to further assume that $w_q[k]$ is white and uncorrelated with *all* samples $e[j]$ up to time $j = k$.
b) In general, β depends on the probability distribution of the quantizer input $e[k]$. This distribution will typically be complicated due to the nonlinear nature of the quantizer that is in feedback with the encoder filter [34]. Since this distribution (and its affect on β) is difficult to characterize as a function of H_{enc}, we assume that β depends only on the number of bits per sample.

These assumptions are widely used and have been proven to be very accurate, particularly at high rate [35].

C. State-Space Signal Model

Substituting the linear quantizer model in (14) into the predictive quantizer system in Fig. 1, one arrives at a linear system depicted in Fig. 2. Also shown in Fig. 2 is a model for the source $z[k]$ in which $z[k]$ is a random noise $w_z[k]$ filtered by a linear filter H_z. This is a standard ARMA model for the source signal. If we assume that $w_z[k]$ is white with zero mean and unit variance, then $z[k]$ will be a wide-sense stationary process with power spectral density $S_z(z) = |H_z(z)|^2$, where $H_z(z)$ is the transfer function of H_z. By appropriately selecting the filter H_z, the model can capture arbitrary second-order statistics of $z[k]$. For example, if we wish to model a lowpass signal $z[k]$, then we can select H_z to be a lowpass filter with the appropriate bandwidth.

For the analysis in this paper, it is convenient to represent H_z in state-space form:

$$\begin{aligned} x[k+1] &= Ax[k] + Bw_z[k], \\ z[k] &= Cx[k] + Dw_z[k], \end{aligned} \tag{16}$$

where $x[k]$ is a vector-valued signal representing the system state, and we assume the noise input $w_z[k]$ is white with zero mean and unit variance. The factors A, B, C, and D are matrices of appropriate dimensions. We will assume the filter H_z is stable with $DD' > 0$. Also, throughout this paper, we will ignore initial conditions and let the time index run from $k = -\infty$ to ∞. Under this assumption, $x[k]$ and $z[k]$ are wide-sense stationary.

State-space modeling for random processes can be found in numerous texts such as [36]–[38]. The presentation in this paper will use the notation adopted in robust control, such as in [39].

D. Kalman Filter Solution for Optimal Closed-Loop Encoder

Having described the various components of the system model, we can now describe the optimal encoder and decoder filters. In the frequency domain, the transfer functions of the optimal filters can be easily derived as Wiener filters [40]. For the LMI analysis in this paper, however, we will need to rederive the formulas in state space using a time-invariant Kalman filter. To the best of our knowledge this state-space solution is new.

Our main interest in the state-space formulas is that they will generalize to the lossy case more easily. However, even in the standard lossless case, the formulas may have some minor benefits. Firstly, we will see that the state-space approach can relatively easily incorporate the effect of the closed-loop quantization noise; frequency-domain analyses of predictive quantization such as in [13], [14] require iterative techniques. Also, the state space solution shows an interesting connection between the effects of quantization and losses on estimation as studied in [15].

Designing the system for a given source $z[k]$ amounts to selecting the encoder and decoder filters, H_{enc} and H_{dec}. For the linear system model, it is also necessary to determine the quantization error variance σ_q^2 that is consistent with the quantizer's coding loss, β. The constraints for this design are specified in the following definition.

Definition 1: Consider the linear predictive filter quantizer in Fig. 2. Fix a stable linear, time-invariant source filter H_z and quantization coding loss $\beta \in [0, 1]$. For this model, an encoder filter H_{enc} will be called *feasible* if

a) H_{enc} is linear time-invariant and strictly causal;
b) the resulting closed-loop map $(w_z, w_q) \mapsto (z, z_E)$ is stable; and
c) there exists a quantization noise level σ_q^2 such that, when the inputs $w_z[k]$ and $w_q[k]$ are zero-mean, unit-variance white noise and $\rho = 1-\beta$, the closed-loop quantizer input signal $z[k] - z_E[k]$ satisfies

$$\sigma_q^2 = \beta(1-\beta)\mathbf{E}|z[k] - z_E[k]|^2. \tag{17}$$

The final constraint (c) requires some explanation. In general, the quantization noise variance is not only a function of the quantizer itself, but also the quantizer input variance that results from the choice of the encoder, H_{enc}. This is a consequence of the fact that the quantization noise scales with the quantizer input variance by the factor β. In Definition 1, the quantization noise level σ_q^2 is thus defined as the level that is consistent with the closed-loop encoder system and the linear quantizer model in Lemma 1. We will call the quantization noise level, σ_q^2 in part (c), the *closed-loop quantization noise variance for the encoder* H_{enc}.

Note that in (17), the expectation does not depend on the time index k. This assumption is valid, since, by the assumptions in parts (a) and (b) of Definition 1, the system is stable and time-invariant. Since the input noise $w[k]$ is wide-sense stationary, all the signals will be wide-sense stationary as well.

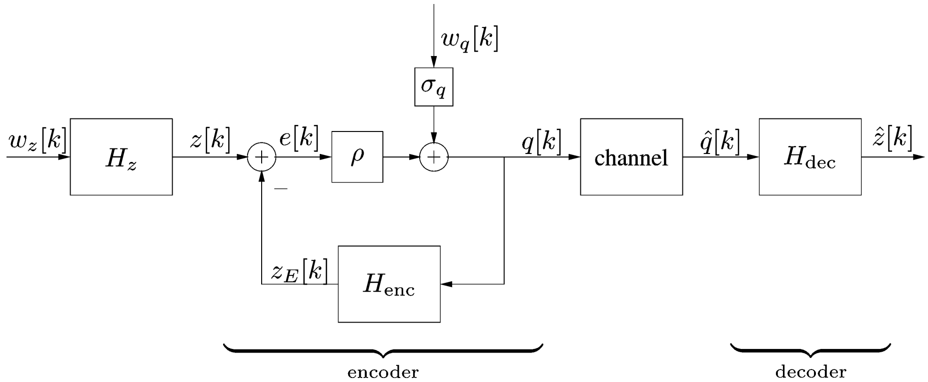

Fig. 3. General predictive quantizer encoder and decoder with a lossy channel. The quantizer is modeled as a linear gain with AWGN noise.

Now, given a feasible encoder $H_{\rm enc}$ and decoder $H_{\rm dec}$, we define the mean-squared error (MSE) of the reconstruction as

$$\mathrm{MSE} = \mathbf{E}|z[k] - z_D[k]|^2 \tag{18}$$

where $z_D[k]$ is the decoder output, and the encoder-decoder system is assumed to operate with the consistent closed-loop quantization noise variance. We will say that a feasible design is *optimal* if it minimizes the MSE.

The following theorem provides a simple state-space characterization of the optimal encoder and decoder filters.

Theorem 2: Consider the linear predictive filter quantizer in Fig. 2, and fix a stable, linear, time-invariant source filter H_z as in (16) and quantization coding loss $\beta \in [0,1]$. Then one optimal encoder filter, $H_{\rm enc}$, can be described by the state space equations

$$\begin{aligned} x_E[k+1] &= Ax_E[k] + Lq[k], \\ z_E[k] &= Cx_E[k] \end{aligned} \tag{19}$$

where the gain matrix L is given by

$$L = E_1 G \tag{20}$$

and E_1 and G are derived from the positive semidefinite solution $P \geq 0$ to the *modified algebraic Riccati equation*

$$\begin{aligned} P &= APA' - (1-\beta)E_1 G E_1' + BB', \\ E_1 &= APC' + BD', \\ G &= (CPC' + DD')^{-1}. \end{aligned} \tag{21}$$

The optimal decoder filter for $H_{\rm dec}$ is given by the state space equations

$$\begin{aligned} x_D[k+1] &= Ax_D[k] + Lq[k], \\ z_D[k] &= Cx_D[k] + q[k]. \end{aligned} \tag{22}$$

Finally, the closed-loop quantization noise variance σ_q^2 and the MSE in (18) are given by

$$\begin{aligned} \sigma_q^2 &= \beta(1-\beta)(CPC' + DD'), \\ \mathrm{MSE} &= \beta(CPC' + DD'). \end{aligned} \tag{23}$$

Proof: See Appendix C-2.

The theorem provides simple state space formulas for the optimal encoder and decoder filters. The state space matrices contain the A and C matrices of the source signal state space system (16), along with a gain matrix L derived from the solution to a matrix (21).

There are two interesting points to note in this result. First, the state space equations for the encoder and decoder filters in (19) and (22), respectively, differ only by a $q[k]$ term in the output equation. Therefore, their transfer functions are related by the simple identity $H_{\rm dec}(z) = H_{\rm enc}(z) + 1$. Moreover, the state update equations in the encoder and decoder equations are identical. Consequently, if the systems have the same initial conditions (i.e., $x_E[0] = x_D[0]$), then $x_E[k] = x_D[k]$ for all subsequent samples k. Thus, the encoder and decoder will have the state estimates. We will see that in the presence of channel losses, this property no longer holds: With losses of the quantizer output, the decoder will not, in general, know the encoder state.

A second point to note is that the modified algebraic Riccati equation in (21) is identical to the main equation in [15] which considers state estimation in the presence of losses of the observed signals. The estimation problem with lossy observations is also considered in [11]. As discussed more precisely in Appendix B, it is shown in [11] and [15] that the MSE in (23) is precisely the coding loss, β, times the one-step ahead prediction error in estimating a signal when the past samples experience independent losses with a probability of β. We thus see an interesting relationship between quantization noise and loss: *the effect of quantization with coding loss β is equivalent to estimating a signal with independent losses with a loss probability of β.* More on the modified algebraic Riccati equation is given in Appendix B.

IV. Predictive Quantization With Losses

A. Markov Erasure Channel Model

We now turn to our main interest: predictive quantization with losses. The scenario we consider is shown in Fig. 3. The system is identical to Fig. 2 except that the encoder output samples, $q[k]$, are transmitted over a lossy channel.

As discussed in Section I, since the quantized samples are uncoded, the effect of losses depends not just on the overall erasure probability, but the exact dynamics of the loss process. We will assume that there is an underlying Markov chain $\theta[k] \in \{0,\ldots,M-1\}$, and the erasures occur when the Markov chain enters a subset of states, $I_{\text{loss}} \subseteq \{0,\ldots,M-1\}$. We define p_{ij} and q_j as in Section II and again assume the Markov chain is aperiodic and irreducible. We will denote the channel output by $\hat{q}[k]$ and take the output to be zero when the sample is lost. Thus,

$$\hat{q}[k] = \begin{cases} q[k] & \text{when } \theta[k] \notin I_{\text{loss}}; \\ 0 & \text{when } \theta[k] \in I_{\text{loss}}. \end{cases} \tag{24}$$

The Markov model is extremely general and captures a large range of erasure processes including independent erasures, Gilbert–Elliot erasures and fixed-length burst erasures.

B. Encoder Design

The goal of robust predictive quantization is: given a signal model (16) and statistics on the Markov erasure channel state θ, to design the encoder H_{enc} and decoder H_{dec} to minimize the average reconstruction error. We first consider the design of the encoder filter H_{enc}.

Following the structure of the optimal encoder (19) for the lossless case, we will assume that the encoder filter for the lossy channel takes the same form:

$$\begin{aligned} x_E[k+1] &= Ax_E[k] + Lq[k], \\ z_E[k] &= Cx_E[k]. \end{aligned} \tag{25}$$

This encoder filter (25) is identical to optimal filter (19) for the lossless system, except that we can use any gain matrix L. In this way, we treat L as a design parameter that can be optimized depending on the loss and source statistics.

It should be stated that, in fixing the encoder to be of the form (25), we have eliminated certain degrees of freedom in the encoder design. The LMI framework would allow the matrices A and C in (25) to differ from the corresponding quantities in (16). But, as we will see, the optimization philosophy in Section IV-E will lead us to (25). We therefore impose this form at the onset to simplify the following discussion.

Our first result for the lossy case characterizes the set of feasible encoder gain matrices.

Theorem 3: Consider the predictive quantizer encoder in Fig. 3 where H_z is described by (16) and the encoder filter H_{enc} is described by (25) for some encoder gain matrix L. Assume that H_z is stable and suppose there exists a $Q \geq 0$ satisfying the Lyapunov equation

$$Q = (A - \rho LC)'Q(A - \rho LC) + C'C \tag{26}$$

with

$$\beta(1-\beta)L'QL < 1. \tag{27}$$

Then the encoder (25) is feasible in the sense of Definition 1, and the corresponding closed-loop quantization noise level is given by

$$\sigma_q^2 = \frac{\beta(1-\beta)}{1-\beta(1-\beta)L'QL} \times \mathbf{Tr}((B-\rho LD)'Q(B-\rho LD) + D'D). \tag{28}$$

Proof: See Appendix C-3.

Theorem 3 provides a simple way of testing the feasibility of a candidate gain matrix L and determining the corresponding quantization noise level. Specifically, the gain matrix L is feasible if the solution Q to the Lyapunov equation (26) satisfies $Q \geq 0$ and the condition in (27). If the gain matrix L is feasible, the closed-loop quantization noise level is given by (28).

It can be shown that the optimal encoder gain L for the lossless system (from Theorem 2) is precisely the gain matrix that minimizes the resulting quantization noise variance σ_q^2. However, in the presence of losses, we will see that the optimal gain does not necessarily minimize the quantization noise: The quantizer input variance can be seen as a measure of how much energy the prediction filter subtracts out from the previous quantizer output samples. With losses, the optimal filter may not subtract out all the energy, thereby leaving some redundancy in the quantizer output samples and thus improving the robustness to losses.

C. Jump Linear Decoder

Having characterized the set of feasible encoders, we can now consider the decoder. Given an encoder, the decoder must essentially estimate the source signal $z[k]$ from the quantizer output samples $q[k]$ that are not lost. We can set this estimation problem up as a jump linear filtering problem and then apply the results in Section II.

To employ the jump linear framework, we need to construct a single jump linear state-space system that describes the signals $q[k]$ and $z[k]$. To this end, we combine the encoder H_{enc}, the source generating filter H_z, and the linear quantizer. The combined system has two states: $x_z[k]$ in the source signal model and $x_E[k]$ in the prediction filter, and we can define the joint state vector $x[k] = [x_z[k]'\ x_E[k]']'$. We also let $w[k]$ denote the joint noise vector $w[k] = [w_z[k]'\ w_q[k]']'$, which contains both the source signal input and quantization noise. We can now view the signals $q[k]$ and $z[k]$ as outputs of a single system whose input is $w[k]$ and state is $x[k]$.

There are two cases for the state and output equations for the system: when the sample $q[k]$ is received by the decoder, and when the sample is lost. We will first consider the case when the sample $q[k]$ is not lost. In this case, $\hat{q}[k] = q[k]$. Using this fact along with (16), (25) and (14), we obtain the larger state space system

$$\begin{aligned} x[k+1] &= A_{\text{NL}}x[k] + B_{\text{NL}}w[k] + u[k], \\ z[k] &= C_1x[k] + D_1w[k], \\ \hat{q}[k] &= C_2x[k] + D_2w[k] \end{aligned} \tag{29}$$

where

$$A_{\text{NL}} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}, \quad B_{\text{NL}} = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$$

$C_1 = [C \quad 0]$, $D_1 = [D \quad 0]$, $C_2 = [\rho C \quad -\rho C]$, $D_2 = [\rho D \quad \sigma_q]$, $u[k] = Uq[k]$, and $U = [0 \quad L']'$. The system (29) expresses the source signal $z[k]$ and the channel output $\hat{q}[k]$ as outputs of a single larger state space system. The system has two inputs: the noise vector $w[k]$ and the signal $u[k]$. When the sample $q[k]$ is not lost over the channel, the input $u[k]$ is known to the decoder. Hence we have added the subscript NL on the matrices, $A_{\rm NL}$ and $B_{\rm NL}$ to indicate the "no loss" matrices.

When the sample $q[k]$ is lost:

$$\begin{aligned} x[k+1] &= A_L x[k] + B_L w[k], \\ z[k] &= C_1 x[k] + D_1 w[k], \\ \hat{q}[k] &= 0 \end{aligned} \tag{30}$$

where $A_{\rm L} = A_{\rm NL} + UC_2$ and $B_{\rm L} = B_{\rm NL} + UD_2$.

Thus, from the perspective of the decoder, the combined signal-encoder system alternates between two possible models: (29) when the sample $q[k]$ is not lost in the channel, and (30) when the sample is lost. Now, in the model we have assumed in Section IV-A, the loss event occurs when the Markov state $\theta[k]$ enters a subset of the discrete states denoted $I_{\rm loss}$. We can thus write the signal-encoder system as a jump linear system driven by the channel Markov state $\theta[k]$: When $\theta[k] = i$,

$$\begin{aligned} x[k+1] &= A_i x[k] + B_i w[k] + u_i[k], \\ z[k] &= C_1 x[k] + D_1 w[k], \\ \hat{q}[k] &= C_{i2} x[k] + D_{i2} w[k] \end{aligned} \tag{31}$$

where the system matrices are given by

$$(A_i, B_i, C_{i2}, D_{i2}) = \begin{cases} (A_{\rm NL}, B_{\rm NL}, C_2, D_2), & \text{when } i \notin I_{\rm loss}; \\ (A_{\rm L}, B_{\rm L}, 0, 0), & \text{when } i \in I_{\rm loss} \end{cases}$$

and $u_i[k]$ is the known input

$$u_i[k] = \begin{cases} U\hat{q}[k] & \text{when } i \notin I_{\rm loss}, \\ 0 & \text{when } i \in I_{\rm loss}. \end{cases}$$

The matrices C_1 and D_1 do not vary with $\theta[k]$.

Following Section II, we can consider a jump linear estimator of the following form: When $\theta[k] = i$,

$$\begin{aligned} \hat{x}[k+1] &= A_i \hat{x}[k] + L_{i1}(\hat{q}[k] - C_{i2}\hat{x}[k]) + u_i[k], \\ \hat{z}[k] &= C_1 x[k] + L_{i2}(\hat{q}[k] - C_{i2}\hat{x}[k]) \end{aligned} \tag{32}$$

where L_{i1} and L_{i2} are gain matrices that are to be determined by the optimization.

Before considering the optimization, it is useful to compare the jump linear estimator (32) with the decoder for the lossless case in (22). The most significant difference is that the decoder in (32) must estimate the states, $x_z[k]$, for the source signal system, as well as the encoder states, $x_E[k]$. In the lossless case, the decoder receives all the quantized samples $q[k]$. Consequently, by running the encoder filter (19), it can reconstruct the encoder state $x_E[k]$. Therefore, the decoder need only estimate the signal state $x_z[k]$. However, with losses, the encoder state is unknown to the decoder and the decoder must estimate both $x_E[k]$ and $x_z[k]$. In particular, if in the lossless problem one must estimate a state of dimension n, the lossy estimator must estimate a state of dimension $2n$.

A second difference is the gain matrices L_{i1} and L_{i2}. In the decoder (22) for the lossless channel, the gain matrices are constant. Moreover, $L_1 = L$ where L is the optimal encoder gain matrix, and $L_2 = I$. In the jump linear decoder (32), the gain matrices L_{i1} and L_{i2} vary with the Markov state $\theta[k]$ and do not necessarily have any simple relation with the encoder matrix.

D. LMI Analysis

Having modeled the system and decoder as jump linear systems, we can now find the optimal decoder gain matrices L_{i1} and L_{i2} in (32) using the LMI analysis in Section II. Specifically, Theorem 1 provides an LMI-constrained optimization for computing the gain matrices L_{i1} and L_{i2} that minimize the MSE,

$$\text{MSE} = \mathbf{E}|z[k] - \hat{z}[k]|^2 \tag{33}$$

where the expectation is over the random signal $z[k]$, the quantization noise $w_q[k]$ and Markov state sequence $\theta[k]$. For the decoder problem, the MSE in (33) is precisely the mean-squared reconstruction error between the original signal $z[k]$ and the decoder output $\hat{z}[k]$. Again, recall that since we have assumed the system is stable and all signals are wide-sense stationary, the MSE in (33) does not depend on the time k.

The LMI method thus provides a simple way of computing the minimum reconstruction error for a given source signal model, predictive encoder and channel loss statistics. By varying the channel loss model parameters, one can thus quantify the effect of channel losses on a predictive quantization system with a given encoder.

The overall analysis algorithm can be described as follows:

Algorithm 1 (LMI Analysis): Consider the predictive quantizer system in Fig. 3. Suppose the source signal $z[k]$ is generated by a stable LTI system of the form (16), the prediction filter can be described by (25) for a given gain matrix L, and the quantizer can be described by (14). Let $\beta \in [0,1]$ be the quantization coding loss, so that $\rho = 1 - \beta$. Suppose the channel erasures can be described by an aperiodic and irreducible M-state Markov chain $\theta[k]$. Then, the optimal decoder of the form (32) can be computed as follows.

1) Use Theorem 3 to determine if the encoder gain matrix L is feasible. Specifically, verify that there is a solution $Q \geq 0$ to the Lyapunov equation (26), and confirm that the solution satisfies (27). If L is not feasible, stop since the MSE is infinite.
2) Compute the closed-loop quantization noise, σ_q^2 in (28).
3) Compute the jump linear system matrices A_i, B_i, C_{i1}, C_{i2}, D_{i1} and D_{i2} as in Section IV-C.
4) Use Theorem 1 to compute the optimal gain matrices L_{i1} and L_{i2} for the decoder (32) and the corresponding reconstruction MSE: $\text{MSE} = \mathbf{E}|z[k] - \hat{z}[k]|^2$.

E. Encoder Gain Optimization

Algorithm 1 in the previous section describes how to compute the minimum MSE achievable at the decoder for a *given* predictive encoder. However, to maximize the robustness of the overall quantizer system, one would like to select the predictive encoder that minimizes this MSE.

To be more specific, let $\text{MSE}(L)$ be the minimum achievable MSE for a given encoder gain matrix L. This minimum error, $\text{MSE}(L)$, can be computed from Algorithm 1 given a model for the signal and a Markov loss model for the channel. Ideally, one would like to search over all gain matrices L to minimize $\mathrm{MSE}(L)$. That is, we wish to compute the optimal encoder gain matrix:

$$L_{\rm opt} = \arg\min_{L} \mathrm{MSE}(L). \tag{34}$$

Unfortunately, this minimization is difficult. The function $\mathrm{MSE}(L)$ is, in general, a complex nonlinear function of the coefficients of L. The global minimum cannot be found without an exhaustive search. If the filter (16) for the signal $z[k]$ has order n, then L will have n coefficients to optimize over. Therefore, even at small filter orders, a direct search over possible gain matrices L will be prohibitively difficult.

To overcome this difficulty, we propose the following simple, but suboptimal, search. For each $\lambda \in [0,1]$, we compute a candidate gain matrix $L(\lambda)$ given by

$$L(\lambda) = E_1(\lambda)G(\lambda) \tag{35}$$

where $E_1(\lambda)$ and $G(\lambda)$ are derived from the solutions to the modified algebraic Riccati equations,

$$\begin{aligned} P(\lambda) &= AP(\lambda)A' - \lambda(1-\beta)E_1GE_1' + BB', \\ E_1(\lambda) &= AP(\lambda)C' + BD', \\ G(\lambda) &= (CP(\lambda)C' + DD')^{-1}. \end{aligned} \tag{36}$$

We can then minimize the MSE, searching over the single parameter λ:

$$L_{\rm subopt} = \arg\min_{\lambda\in[0,1]} \mathrm{MSE}(L(\lambda)). \tag{37}$$

Similarly to (34), the minimization in (37) is not necessarily convex, and an exact solution would require an exhaustive search. However, since λ is a scalar parameter and dependence on λ is continuous, a good approximate solution to (37) can be found by testing a small number of values of $\lambda \in [0,1]$.

This suboptimal search over the candidate gain matrices $L(\lambda)$ can be motivated as follows: From Theorem 2, we see that when $\lambda = 1$, $L(\lambda)$ is precisely the optimal encoder gain matrix for the lossless system. For other values of λ, $L(\lambda)$ is the optimal gain matrix for a lossless channel, but with a higher effective coding loss given by $\beta_{\rm eff} = 1 - \lambda(1-\beta)$. As λ decreases from 1 to 0, this effective coding loss, $\beta_{\rm eff}$, increases from β to 1. The increase in the coding loss $\beta_{\rm eff}$ represents an increase in the effective quantization noise. The logic in the suboptimal search (37) is that the modifications for the encoder for higher quantization noise should be qualitatively similar to the modifications necessary for channel losses. As the quantization noise increases, the prediction filter is less able to rely on past quantized samples and will naturally decrease the weighting of those samples in the prediction output. This removal of past samples from the prediction output will leave redundancy in the quantized samples and should improve the robustness to channel losses.

V. Numerical Examples

To illustrate the robust predictive quantization design method, we consider the quantization of the output of a second-order Chebyshev lowpass filter with cutoff frequency $\pi/10$ driven by a zero-mean white Gaussian input. For the quantizer $Q(\cdot)$, we use an optimally-loaded uniform quantizer with 2 bits per sample.

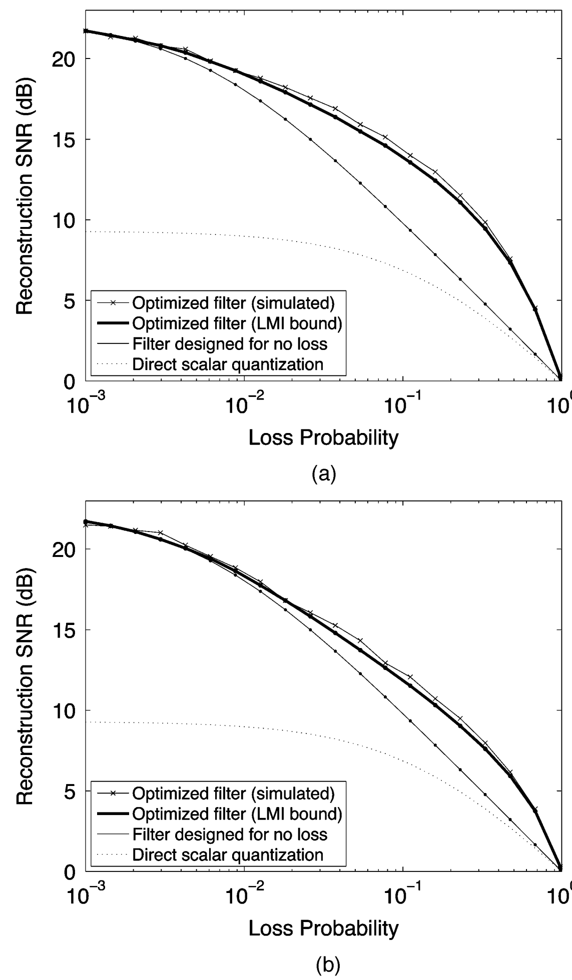

Fig. 4. Numerical examples of robust predictive quantization of a lowpass signal with two channel types. For each channel type, the reconstruction SNR as a function of channel loss probability is plotted for three quantizers: (i) scalar quantization with no interpolation of lost samples at the decoder; (ii) the optimal predictive encoder and decoder assuming no channel losses; and (iii) the proposed robust predictive quantizer where the encoder and decoder are optimized for each loss probability. (a) i.i.d. losses and (b) Gilbert–Elliot channel.

Fig. 4(a) shows the reconstruction signal-to-noise ratio (SNR) for various systems. In each case, the reconstruction SNR is plotted as a function of the channel loss probability assuming the channel losses are i.i.d. The reconstruction SNR in dB is defined as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\mathbf{E}|z[k]|^2}{\mathbf{E}|z[k]-\hat{z}[k]|^2}\right)\mathrm{dB}$$

where $z[k]$ is the lowpass signal to be quantized and $\hat{z}[k]$ is the final estimate at the decoder.

The bottom curve in Fig. 4(a) is the reconstruction SNR with direct scalar quantization at the encoder and no interpolation for lost samples at the decoder. That is, the decoder simply substitutes a value of zero for any lost samples. When there are no losses, scalar quantization achieves a reconstruction SNR of 9.3 dB, which is the coding gain for an optimal 2-bit uniform quantizer. Since there is no interpolation at the decoder, with losses, the reconstruction error decays linearly with the loss probability.

The second curve in Fig. 4(a) is obtained with the optimal encoder designed for no losses as described in Section III-D. With no losses, this encoder achieves a reconstruction SNR of approximately 21.7 dB which represents a prediction gain of approximately $21.7 - 9.3 = 12.4$ dB relative to scalar quantization. However, in the presence of losses the improvement due to prediction rapidly decreases. For example, at a loss probability of 20%, the predictive quantizer performs less than 2 dB better than scalar quantization with no prediction. In this sense, the predictive quantizer which is not designed for losses is not robust.

The bold curve depicts the computed performance of robust predictive quantization when the encoder is optimized for the loss probability. Specifically, for each loss probability, the minimization (37) is performed to find the optimal encoder gain matrix, and the MSE at the decoder is based on the optimal jump linear decoder described in Section IV-C. For the line-search minimization in (37), we adaptively selected 20 values of λ for each loss probability. By comparing to a more dense sampling of λ for a few loss probabilities, we concluded that this method was adequate to get within 0.05 dB of the optimal performance.

Finally, the $\times$'s connected by a thin line show simulated performance using the optimized prediction filters described above, 2-bit uniform scalar quantization, and the linear time-varying Kalman filter estimator. The simulated performance tracks the LMI-based computations closely. The small gap is primarily due to the fact that the LMI bound assumes a suboptimal jump linear estimator, while the simulated performance is based on the optimal time-varying Kalman filter.

With no losses, the robust predictive quantizer is identical to the optimal encoder with no losses. As the loss probability increases, the degradation of the robust predictive quantizer is much smaller than the degradation of the encoder optimized for no loss. For example, at a 20% loss probability, the robust predictive quantizer has a reconstruction SNR more than 4.5 dB greater than the system without encoder optimization.

The same design method can be used for more complicated channels. Fig. 4(b) repeats the previous example with a Gilbert–Elliot loss model—a simple channel with correlated losses. The good-to-bad state transition parameter λ_1 is varied from 0 to 1, and the bad-to-good parameter is set to $\lambda_2 = (1-\lambda_1)/2$. The resulting loss probability is $2\lambda_1/(1+\lambda_1)$. As can be seen in Fig. 4(b), the robust predictive quantizer again shows an improved performance over the encoder optimized for no losses. As with Fig. 4(a), the results computed via convex optimizations are supported by Monte Carlo simulation.

Note well that we are comparing techniques that do not induce any delay. When arbitrarily large delay is allowed, considerably better performance is possible. For example, consider i.i.d. losses with probability 0.5. Since the example uses 2 bits per sample, one can achieve a reliable rate of 1 bit per sample using ideal forward error correction (FEC). It can be verified using the methods in Section III-D that such a system would achieve a reconstruction SNR of 12.4 dB for the source under consideration. In comparison, the optimized robust predictive quantizer achieves a much lower SNR of only 7.0 dB.

Along with the issue of delay, another problem with conventional block FEC is the necessity to match the code rate with the probability of loss: if too few codewords are received, decoding is impossible and performance is poor; and additional codewords beyond the minimum needed to decode do not help. The proposed predictive quantizer design method is robust over a range of channel conditions. To illustrate this, Fig. 5 shows the performance of a predictive coder with the filter that is optimized for a fixed loss rate of 10%, but evaluated over a range of actual losses rates from 0% to 100%. At each actual loss probability, the performance of the predictive quantizer optimized for the 10% loss is compared against the encoder optimized for the true loss. It can be seen that the predictive quantizer optimized for the 10% loss performs very closely to the optimal over a range of loss probabilities. Specifically, the filter designed for 10% loss has a reconstruction error within 0.4 dB of the optimal filter for all loss rates above 5%. Meanwhile, the performance with a filter designed for no loss degrades quickly in the presence of loss. The conclusion is that a filter designed for a small loss rate can perform well over a range of loss conditions.

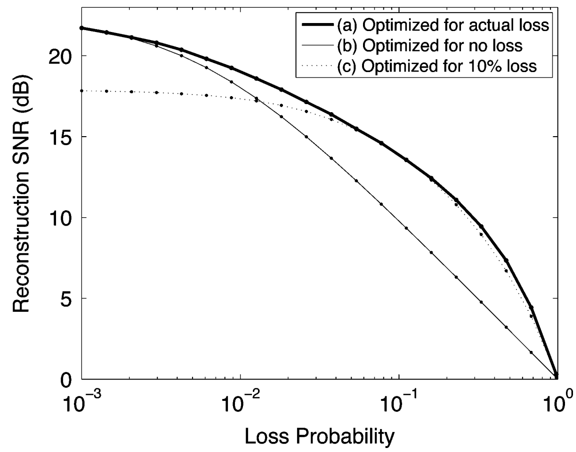

Fig. 5. Quantizer performance as a function the loss probability with three different prediction filters: (a) robust prediction filter designed for the actual loss probability; (b) filter optimized for no loss; and (c) filter optimized for 10% loss. Simulation conditions are identical to those in Fig. 4(b).

VI. Concluding Comments

While the effect of losses in predictive quantization is well known at a qualitative level, there have been few practical methods to quantify the effects of such losses. This paper has presented a novel method based on jump linear system modeling and LMI-constrained convex optimization that can be used to precisely compute the overall degradation in reconstruction from the signal model and loss statistics. The method applies to arbitrary-order prediction filters and general Markov erasure channels. Other effects such as the delay in the reconstruction can also easily be incorporated.

Also presented is a method for modifying the prediction filter to improve the robustness of the encoder. The approach, based on leaving redundancy in the signal, illustrates a general tradeoff between compression and robustness. The proposed method searches over a small but intuitively-reasonable class of encoders. While our simulations indicate that these encoders can significantly improve the robustness to losses, they are not necessarily optimal. We also have not explored varying the sampling rate, e.g., to provide opportunity for channel coding.

APPENDIX A
STEADY-STATE KALMAN FILTER EQUATIONS

The Kalman filter is widely used in linear estimation problems and covered in several standard texts such as [31] and [38]. However, the analysis in this paper requires a slightly non-standard form, used in modern robust control texts such as [39]. It is useful to briefly review the equations that will be used.

To describe the Kalman filter, consider a linear state space system of the form

$$\begin{aligned} x[k+1] &= Ax[k] + Bw[k], \\ z[k] &= C_1x[k] + D_1w[k], \\ y[k] &= C_2x[k] + D_2w[k]. \end{aligned} \tag{38}$$

We assume the input $w[k]$ is a white Gaussian process with zero mean and unit variance. The Kalman filtering problem is to estimate the first output signal $z[k]$ and state $x[k]$ from the second output $y[k]$. The signal $z[k]$ represents an unknown signal that we wish to estimate, and $y[k]$ is the observed signal.

Define the state and output estimates:

$$\begin{aligned} \hat{x}[k|j] &= \mathbf{E}(x[k]|y[j], y[j-1], \ldots), \\ \hat{z}[k|j] &= \mathbf{E}(z[k]|y[j], y[j-1], \ldots). \end{aligned}$$

Thus, $\hat{x}[k|j]$ is the MMSE estimate of the state $x[k]$ given the observed output up to sample j. While the Kalman filter can be used to compute the estimates for any k and j, we will be interested here in three specific estimates:

- $\hat{z}[k|k]$: The *causal* estimate of $z[k]$ given the observations of $y[j]$ up to time $j = k$.
- $\hat{z}[k|k-1]$: The *strictly causal* estimate of $z[k]$ given the observations of $y[j]$ up to time $j = k-1$.
- $\hat{x}[k|k-1]$: The strictly causal estimate of the state $x[k]$.

To avoid the effect of initial conditions, we will assume all signals are wide-sense stationary. Under the additional technical assumptions that (A, C_2) is detectable and $D_2D_2' > 0$, it can be shown that the optimal estimate for the above three quantities is given by recursive *Kalman filter* equations:

$$\begin{aligned} \hat{x}[k+1|k] &= A\hat{x}[k|k-1] + L_1(y[k] - C_2\hat{x}[k|k-1]) \\ \hat{z}[k|k-1] &= C_1\hat{x}[k|k-1] \\ \hat{z}[k|k] &= C_1\hat{x}[k|k-1] + L_2(y[k] - C_2\hat{x}[k|k-1]) \end{aligned} \tag{39}$$

where L_1 and L_2 are known as the *Kalman gain matrices* and will be described momentarily. We see that the Kalman filter is itself a linear time-invariant state-space filter whose input is $y[k]$ and outputs are the estimates of $x[k]$ and $z[k]$.

The filter (39) are sometimes called the *time-invariant* or *steady-state* equations, since the gain matrices L_1 and L_2 do not vary with time. In general, to account for either initial conditions or time-varying systems, the gain matrices L_1 and L_2 would need to be time-varying. However, since we will only be interested in time-invariant systems and asymptotic performance, the time-invariant form considered here is sufficient.

The gain matrices L_1 and L_2 in (39) can be determined from the solution to the well-known *algebraic Riccati equation* given by

$$P = APA' + BB' - E_1GE_1' \tag{40}$$

where $E_1 = APC_2' + BD_2'$ and $G = (C_2PC_2' + D_2D_2')^{-1}$. Under the assumption that (A, C_2) is detectable and $D_2D_2' > 0$, it can be shown that (40) has a unique positive semi-definite solution P. The gain matrices are then given by

$$L_1 = E_1G, \quad L_2 = E_2G \tag{41}$$

where

$$E_2 = C_1PC_2' + D_1D_2'. \tag{42}$$

The algebraic Riccati equation solution P has the interpretation of the state-estimation error variance:

$$P = \mathbf{E}((x[k] - \hat{x}[k|k-1])((x[k] - \hat{x}[k|k-1])').$$

Also, the mean-squared output errors are given by

$$\sigma^2[k|k-1] = \mathbf{Tr}(C_1PC_1' + D_1D_1'), \tag{43}$$

$$\sigma^2[k|k] = \mathbf{Tr}(C_1PC_1' + D_1D_1' - E_2GE_2) \tag{44}$$

where $\sigma^2[k|j] = \mathbf{E}\|z[k] - \hat{z}[k|j]\|^2$.

APPENDIX B
MODIFIED ALGEBRAIC RICCATI EQUATION

The analysis in Section III-D results in a modified form of the standard algebraic Riccati equation,

$$\begin{aligned} P &= APA' - (1-\beta)E_1GE_1' + BB', \\ E_1 &= APC' + BD', \\ G &= (CPC' + DD')^{-1} \end{aligned} \tag{45}$$

for a loss parameter $\beta \in [0, 1]$. When $\beta = 0$, the equation reduces to the standard algebraic Riccati equation (40); when $\beta = 1$, it reduces to the discrete Lyapunov equation $P = APA' + BB'$.

A similar equation arises in [15] and [11], which consider state estimation of a system (38) with i.i.d. erasures of the observation signal $y[k]$. A result in [11] specifically shows the following: Suppose that $y[k]$ is the output of a linear state-space system

$$\begin{aligned} x[k+1] &= Ax[k] + Bw[k], \\ z[k] &= Cx[k] + Dw[k] \end{aligned} \tag{46}$$

where $w[k]$ is unit-variance white noise. Let $\hat{z}[k|k-1]$ represent the one-step ahead prediction of $z[k]$ from past samples $y[k-1]$, $y[k-2], \ldots$, where

$$y[k] = \begin{cases} z[k], & \text{with probability } 1-\beta; \\ 0, & \text{with probability } \beta. \end{cases}$$

That is, $\hat{z}[k|k-1]$ is the estimate of $z[k]$ from past samples, where the samples are lost with probability β. If the losses are i.i.d. and $\hat{z}[k|k-1]$ is the optimal jump linear estimator (discussed in Section II), it is shown in [11] that the resulting reconstruction error is given by

$$\mathbf{E}\|z[k]-\hat{z}[k|k-1]\|^2 = \mathbf{Tr}(CPC'+DD')$$

where P is the solution to the modified algebraic Riccati equation (45). Comparing this result with Theorem 2, we see that there is one-to-one correspondence between the effects of i.i.d. losses and quantization.

As discussed in [11], the modified algebraic Riccati equation can be solved via an LMI-constrained optimization. However, in the special case of a scalar output, the following simple iterative procedure is also possible.

Suppose D has a scalar output, i.e., D is a row vector. For $\sigma > 0$, define the matrix function $D(\sigma) = [D\,\sigma]$. Then, it is easy to verify that P is a solution to (45) if and only if there exists a $\sigma > 0$ satisfying the following coupled equations:

$$P = APA' - E_1(CPC'+D(\sigma)D(\sigma)')^{-1}E_1' + BB' \quad (47)$$

$$E_1 = APC' + BD' \quad (48)$$

$$\sigma^2 = \frac{\beta}{1-\beta}(CPC'+DD'). \quad (49)$$

For a fixed σ, (47)–(48) is a standard algebraic Riccati equation and can be easily solved for P. Also, it can be verified that the solution P monotonically increases with σ. This leads to the following iterative bisection search procedure.

- *Step 1:* Find minimum and maximum values, $\sigma_{\min}$ and $\sigma_{\max}$, for the possible solution σ.
- *Step 2:* Set σ to the midpoint of the interval, $\sigma = (\sigma_{\min} + \sigma_{\max})/2$.
- *Step 3:* Solve the algebraic Riccati equation (47)–(48) for P using the midpoint value of σ.
- *Step 4:* If $\sigma^2 > (\beta/(1-\beta))(CPC'+DD')$ then set $\sigma_{\max} = \sigma$ and return to Step 2; otherwise, set $\sigma_{\min} = \sigma$ and return to Step 2.

This iterative procedure will converge exponentially to a (σ, P) pair satisfying the coupled (47)–(49).

Appendix C
Proofs

1) Proof of Lemma 1

To simplify the notation in this proof, we will omit the time index k on all signals. From (12), $\mathbf{E}(qe\,|\,q) = q^2$, so

$$\mathbf{E}(qe) = \mathbf{E}(q^2). \quad (50)$$

Combining (50) and (13), we obtain

$$\beta\mathbf{E}e^2 = \mathbf{E}(q-e)^2 = \mathbf{E}q^2 - 2\mathbf{E}qe + \mathbf{E}e^2 = -\mathbf{E}q^2 + \mathbf{E}e^2.$$

Therefore,

$$\mathbf{E}q^2 = (1-\beta)\mathbf{E}e^2. \quad (51)$$

Now, let $v = q-(1-\beta)e$, so that $q = (1-\beta)e + v$. Since q and e are zero mean, so is v. Also, using (50) and (51), $\mathbf{E}(ve) = \mathbf{E}(q-(1-\beta)e)e) = (1-\beta)\mathbf{E}(e^2) - (1-\beta)\mathbf{E}(e^2) = 0$, so v and e are uncorrelated. Finally,

$$\begin{aligned}\mathbf{E}(v^2) &= \mathbf{E}(q-(1-\beta)e)^2\\ &= \mathbf{E}(q^2) - 2(1-\beta)\mathbf{E}(qe) + (1-\beta)^2\mathbf{E}(e^2)\\ &= [(1-\beta) - 2(1-\beta)^2 + (1-\beta)^2]\mathbf{E}(e^2)\\ &= \beta(1-\beta)\mathbf{E}(e^2).\end{aligned}$$

Therefore, if we define σ_q and ρ as in (15) and let $w_q = (1/\sigma_q)v$, we have that w_q is a zero-mean process, uncorrelated with e, with unit variance and satisfying $q = \rho e + \sigma_q w_q$. □

2) Proof of Theorem 2

We know from [1] that the optimal encoder and decoder filters must be the MMSE estimators,

$$\begin{aligned}z_E[k] &= \hat{z}[k\,|\,k-1] = \mathbf{E}(z[k]\,|\,q[0:k-1]),\\ z_D[k] &= \hat{z}[k\,|\,k] = \mathbf{E}(z[k]\,|\,q[0:k]).\end{aligned}$$

That is, $z_E[k]$ must be the MMSE estimate of $z[k]$ given the quantized samples $q[j]$ up to time $j = k-1$. The optimal decoder output is the MMSE estimate using the samples up to time $j = k$. Since the system is linear and time-invariant, both estimators are given by standard time-invariant Kalman filters.

The Kalman filter equations are summarized in Appendix A. To employ the equations, we need to first combine the quantizer and signal model into a single state-space system. To this end, we combine (16) with (14) to obtain

$$\begin{aligned}x[k+1] &= Ax[k] + Bw_z[k],\\ z[k] &= Cx[k] + Dw_z[k],\\ q[k] &= \rho Cx[k] + \rho Dw_z[k] + \sigma_q w_q[k] + u[k]\end{aligned} \quad (52)$$

where $u[k] = -\rho z_E[k]$. If we define the vector noise $w[k] = [w_z[k]\ \ w_q[k]]$ and let $B_1 = [B\ \ 0]$, $C_1 = C$, $C_2 = \rho C$, $D_1 = [D\ \ 0]$, and $D_2 = [\rho D\ \ \sigma]$, then (52) can be rewritten as

$$\begin{aligned}x[k+1] &= Ax[k] + B_1 w[k],\\ z[k] &= C_1 x[k] + D_1 w[k],\\ q[k] &= C_2 x[k] + D_2 w[k] + u[k].\end{aligned} \quad (53)$$

We have now described the observed output $q[k]$ and signal to be estimated $z[k]$ as outputs of a single linear state-space system. We can then apply standard Kalman filter equations from Appendix A to obtain the estimator

$$\begin{aligned}\hat{x}[k+1\,|\,k] &= A\hat{x}[k\,|\,k-1]\\ &\quad + L(q[k] - C_2\hat{x}[k\,|\,k-1] - u[k]),\\ \hat{z}[k\,|\,k-1] &= C_1\hat{x}[k\,|\,k-1]\end{aligned}$$

where L is the Kalman gain matrix. Note that we have used the fact that $u[k]$ is a known input, since it can be computed from $z_E[k] = \hat{z}[k|k-1]$. If we define the state $x_E[k] = \hat{x}[k|k-1]$, and use the fact that $z_E[k] = \hat{z}[k|k-1]$, the encoder equations can be rewritten as

$$\begin{aligned} x_E[k] &= Ax_E[k] + L(q[k] - C_2 x_E[k] - u[k]), \\ z_E[k] &= C_1 x_E[k]. \end{aligned} \tag{54}$$

Now, since $C_2 = \rho C$,

$$u[k] = -\rho z_E[k] = -\rho C\hat{x}[k|k-1] = -C_2\hat{x}[k|k-1]. \tag{55}$$

Using this identity along with the fact that $C_1 = C$ in (54), we obtain the encoder equations in (19).

Next, we derive the expression for the Kalman gain matrix L. To this end, first observe that, using (44), the quantizer input variance is given by

$$\mathbf{E}\|e[k]\|^2 = \mathbf{E}\|z[k] - \hat{z}[k \mid k-1]\|^2 = CPC' + DD',$$

where P is the error variance matrix. Using (17), the quantizer error variance is given by

$$\sigma_q^2 = \beta(1-\beta)\mathbf{E}\|e[k]\|^2 = \beta(1-\beta)(CPC' + DD').$$

Now using the fact that $C_2 = \rho C$ and $D_2 = [\rho D \sigma_q]$, the matrix G in (40) is given by

$$\begin{aligned} G^{-1} &= C_2PC_2' + D_2D_2' = \rho^2 CPC' + \rho^2 DD' + \sigma_q^2 \\ &= (1-\beta)^2(CPC' + DD') + \beta(1-\beta)(CPC' + DD') \\ &= (1-\beta)(CPC' + DD'). \end{aligned} \tag{56}$$

Thus,

$$\begin{aligned} P &= APA' + (APC_2' + B_1D_2')G(C_2PA' + D_2B_1') + B_1B_1' \\ &= APA' + \frac{\rho^2}{1-\beta}APC'(CPC' + DD')^{-1}CPA' + BB' \\ &= APA' + (1-\beta)(APC' + BD') \\ &= (CPC' + DD')^{-1}(CPA' + DB') + BB' \\ &= APA' + (1-\beta)E_1GE_1' + BB' \end{aligned}$$

where $E_1 = APC' + BD'$ and $G = (CPC' + DD')^{-1}$. This proves (21). Substituting (56) into the expression for L_1 in (41):

$$\begin{aligned} L = L_1 &= (APC_2' + B_1D_2')G \\ &= \frac{\rho}{1-\beta}(APC' + BD')(CPC' + DD')^{-1} \\ &= (APC' + BD')(CPC' + DD')^{-1} = E_1G \end{aligned}$$

which proves (20).

Now, substituting the expressions for C_1, C_2, D_1 and D_2 into the expression for E_2 in (42), we obtain

$$E_2 = C_1PC_2' + D_1D_2' = \rho(CPC' + DD'). \tag{57}$$

Substituting (55), (56) and (57) in (39),

$$\begin{aligned} \hat{z}[k|k-1] &= C_1\hat{x}[k|k-1] = C\hat{x}[k|k-1], \\ \hat{z}[k|k] &= C_1\hat{x}[k|k-1] \\ &\quad + EG(q[k] - C_2\hat{x}[k|k-1] - u[k]) \\ &= C\hat{x}[k|k-1] + q[k]. \end{aligned}$$

As discussed above, the optimal decoder output is given by $z_D[k] = \hat{z}[k \mid k]$. Therefore, if we define the decoder state as $x_D[k] = \hat{x}[k \mid k-1]$, we obtain the decoder equations

$$\begin{aligned} x_D[k] &= Ax_D[k] + Lq[k], \\ z_D[k] &= Cx_D[k] + q[k]. \end{aligned} \tag{58}$$

Finally, substituting (56) and (57) into the expression for the MSE in (43),

$$\begin{aligned} \text{MSE} &= \mathbf{E}|z[k] - z_D[k]|^2 = \mathbf{E}|z[k] - \hat{z}[k \mid k]|^2 \\ &= C_1PC_1' + D_1D_1' - E_2GE_2' \\ &= CPC' + DD' - \frac{\rho^2}{1-\beta}(CPC' + DD') \\ &= \beta(CPC' + DD'). \end{aligned}$$

3) *Proof of Theorem 3*

We need to show that, if L satisfies the conditions in Theorem 3, the resulting encoder in (25) satisfies the three conditions in Definition 1. For any gain matrix L, the encoder (25) is linear, time invariant and strictly causal, and therefore satisfies condition (a) of Definition 1.

To prove condition (b), define the error signals, $e_x[k] = x_z[k] - x_E[k]$ and $e_z[k] = z[k] - z_E[k]$. Combining (16), (25) and (14), we obtain

$$\begin{aligned} e_x[k+1] &= (A - \rho LC)e_x[k] \\ &\quad + (B - \rho LD)w_z[k] - \sigma_q Lw_q[k], \\ e_z[k] &= Ce_z[k] + Dw_z[k]. \end{aligned} \tag{59}$$

The system (59) is a standard LTI system. If we define the closed-loop matrices $A_{\rm L} = A - \rho LC$ and

$$B_{\rm L} = \begin{bmatrix} B - \rho LD \\ \sigma_q L \end{bmatrix}$$

and let $w[k] = [w_z[k]' \; w_q[k]']'$, then $w[k]$ is unit-variance white noise, and (59) can be rewritten as

$$\begin{aligned} e_x[k+1] &= A_{\rm L}e_x[k] + B_{\rm L}w[k], \\ e_z[k] &= Ce_z[k] + Dw_z[k]. \end{aligned} \tag{60}$$

Now, suppose there exists a $Q \geq 0$ satisfying (26). The condition is equivalent to

$$Q = A_{\rm L}'QA_{\rm L} + C'C. \tag{61}$$

Now, since H_z is stable, A is stable. Also, since $A_{\rm L} = A - \rho LC$, for any matrix C and L, $(A_{\rm L}, C)$ is detectable. By a standard result for Lyapunov equations (see, for example, [37]), the existence of a matrix $Q \geq 0$ satisfying (61) implies that $A_{\rm L}$ is stable and

$$\begin{aligned} \mathbf{E}|e_z[k]|^2 &= \mathbf{Tr}(B_{\rm L}'QB_{\rm L} + D'D) \\ &= \mathbf{Tr}((B - \rho LD)'Q(B - \rho LD) \\ &\quad + D'D + \sigma_q^2 L'QL). \end{aligned} \tag{62}$$

This in turn implies that, for any σ_q, the mapping $(w_z, w_q) \mapsto e_x$ is stable. Also, since A is stable, the map $w_z \mapsto x$ is stable. Therefore, since $e_x = x - x_E$, the mapping $(w_z, w_q) \mapsto (x, x_E)$ is well-posed and stable and the gain matrix L satisfies condition (b).

For condition (c), we must show that if σ_q is defined as in (28), the resulting closed-loop system satisfies $\sigma_q^2 = \beta(1-\beta)\sigma_z^2$, where $\sigma_z^2 = \mathbf{E}|z[k] - z_E[k]|^2 = \mathbf{E}|e_z[k]|^2$. Combining (62) and (28),

$$\begin{aligned}\sigma_q^2 &- \beta(1-\beta)\sigma_z^2\\ &= \sigma_q^2 - \beta(1-\beta)\mathbf{Tr}((B-\rho LD)'Q(B-\rho LD)\\ &\quad + D'D + \sigma_q^2 L'QL)\\ &= (1-\beta(1-\beta)L'QL)\sigma_q^2\\ &\quad - \beta(1-\beta)\mathbf{Tr}((B-\rho LD)'Q(B-\rho LD) + D'D)\\ &= 0.\end{aligned}$$

Hence, $\sigma_q^2 = \beta(1-\beta)\sigma_z^2$ and, thus, σ_q^2 is the quantization noise level for L. □

REFERENCES

[1] A. Gersho and R. M. Gray*, Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

[2] R. V. Cox, "Speech coding," in *The Digital Signal Processing Handbook*. Boca Raton, FL: CRC and IEEE Press, 1998, ch. 45, pp. 45.1–45.19.

[3] A. Gersho, "Advances in speech and audio compression," *Proc. IEEE*, vol. 82, no. 6, pp. 900–918, Jun. 1994.

[4] D. LeGall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, no. 4, pp. 46–58, Apr. 1991.

[5] T. M. Cover and J. A. Thomas*, Elements of Information Theory*. New York: Wiley, 1991.

[6] M. Luby, "LT codes," in *Proc. IEEE Symp. Found. Comp. Sci.*, Vancouver, BC, Canada, Nov. 2002, pp. 271–280.

[7] M. Ait Rami and L. El Ghaoui, "LMI optimization for nonstandard Riccati equations arising in stochastic control," *IEEE Trans. Automat. Control*, vol. 41, no. 11, pp. 1666–1671, Nov. 1996.

[8] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan*, Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.

[9] O. L. V. Costa, M. D. Fragoso, and R. P. Marques*, Discrete-Time Markov Jump Linear Systems*, ser. Probability and Its Applications. London, U.K.: Springer, 2005.

[10] O. L. V. Costa, J. B. R. do Val, and J. C. Geromel, "A convex programming approach to H_2-control of discrete-time Markovian jump linear systems," *Int. J. Control*, vol. 66, no. 4, pp. 557–559, Apr. 1997.

[11] A. K. Fletcher, "A jump linear framework for estimation and robust communication with Markovian source and channel dynamics," Ph.D. dissertation, Univ. California, Berkeley, 2005.

[12] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Causal and strictly causal estimation for jump linear systems: An LMI analysis," in *Proc. Conf. Inform. Sci. & Sys.*, Princeton, NJ, Mar. 2006.

[13] E. G. Kimme and F. F. Kuo, "Synthesis of optimal filters for a feedback quantization scheme," *IEEE Trans. Circuits Syst.*, vol. 10, no. 3, pp. 405–413, Sep. 1963.

[14] O. G. Guleryuz and M. T. Orchard, "On the DPCM compression of Gaussian autoregressive sequences," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 945–956, Mar. 2001.

[15] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1453–1464, Sept. 2004.

[16] Y.-L. Chen and B.-S. Chen, "Model-based multirate representation of speech signals and its applications to recover of missing speech packets," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 220–231, May 1997.

[17] E. Gündüzhan and K. Momtahan, "A linear prediction based packet loss concealment algorithm for PCM coded speech," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 778–785, Nov. 2001.

[18] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[19] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 61–82, July 2000.

[20] W.-N. Lie and Z.-W. Gao, "Video error concealment by integrating greedy suboptimization and Kalman filtering techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 8, pp. 982–992, Aug. 2006.

[21] R. Anand, K. Ramchandran, and I. V. Kozintsev, "Continuous error detection (CED) for reliable communication," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1540–1549, Sept. 2001.

[22] G. Buch, F. Burkert, J. Hagenauer, and B. Kukla, "To compress or not to compress?," in *IEEE Globecom*, London, Nov. 1996, pp. 196–203.

[23] J. Chou and K. Ramchandran, "Arithmetic coding-based continuous error detection for efficient ARQ-based image transmission," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 861–867, June 2000.

[24] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2449–2457, Sept. 1995.

[25] K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Trans. Comm.*, vol. 39, no. 6, pp. 838–846, Jun. 1991.

[26] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 48–57, Jan. 1995.

[27] D. L. Neuhoff, "On the asymptotic distribution of errors in vector quantization," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 461–468, Mar. 1996.

[28] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Automat. Control*, vol. 33, no. 8, pp. 780–783, Aug. 1988.

[29] O. L. V. Costa, "Linear minimum mean square error estimation for discrete-time Markovian jump linear systems," *IEEE Trans. Automat. Control*, vol. 39, no. 8, pp. 1685–1689, Aug. 1994.

[30] A. Doucet, A. Logothetis, and V. Krishnamurthy, "Stochastic sampling algorithms for state estimation of jump Markov linear systems," *IEEE Trans. Automat. Control*, vol. 45, no. 1, pp. 188–202, Jan. 2000.

[31] A. V. Balakrishnan*, Kalman Filtering Theory*. New York: Springer, Feb. 1984.

[32] J. A. Bucklew and N. C. Gallagher, Jr, "Some properties of uniform step size quantizers," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 5, pp. 610–613, Sept. 1980.

[33] H. Gish and J. P. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, no. 5, pp. 676–683, Sept. 1968.

[34] R. M. Gray, "Quantization noise spectra," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1220–1244, Nov. 1990.

[35] H. Viswanathan and R. Zamir, "On the whiteness of high-resolution quantization errors," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 2029–2038, July 2001.

[36] P. E. Caines*, Linear Stochastic Systems*. New York: Wiley, 1988.

[37] F. M. Callier and C. A. Desoer*, Linear Systems Theory*. New York: Springer, 1994.

[38] P. R. Kumar and P. Varaiya*, Stochastic Systems: Estimation, Identification, and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[39] K. Zhou, J. C. Doyle, and K. Glover*, Robust and Optimal Control*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[40] P. M. Clarkson*, Optimal and Adaptive Signal Processing*. Boca Raton, FL: CRC, 1993.

Alyson K. Fletcher (M'07) received the B.S. degree in mathematics from the University of Iowa, Iowa City, and the M.S. degree in electrical engineering in 2002 and the M.A. degree in mathematics and Ph.D. degree in electrical engineering, both in 2006, from the University of California, Berkeley.

She is currently a President's Postdoctoral Fellow at the University of California, Berkeley. Her research interests include estimation, image processing, statistical signal processing, sparse approximation, wavelets, and control theory.

Dr. Fletcher is a member of SWE, SIAM, and Sigma Xi. In 2005, she received the University of California Eugene L. Lawler Award, the Henry Luce Foundation's Clare Boothe Luce Fellowship, and the Soroptimist Dissertation Fellowship.

Sundeep Rangan received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley in 1995 and 1997, respectively.

He was then a Postdoctoral Research Fellow at the University of Michigan, Ann Arbor. He joined the Wireless Research Center at Bell Laboratories, Lucent Technologies, in 1998. In 2000, he co-founded, with four others, Flarion Technologies which developed and commercialized an early OFDM cellular wireless data system. Flarion was acquired by Qualcomm in 2006. He is currently a Director of Engineering at Qualcomm Technologies, where he is involved in the development of next generation cellular wireless systems. His research interests include communications, wireless systems, information theory, estimation, and control theory.

Vivek K Goyal (S'92–M'98–SM'03) received the B.S. degree in mathematics and the B.S.E. degree in electrical engineering (both with highest distinction) from the University of Iowa, Iowa City, in 1993. He received the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1995 and 1998, respectively.

He was a Member of Technical Staff in the Mathematics of Communications Research Department of Bell Laboratories, Lucent Technologies, 1998–2001; and a Senior Research Engineer for Digital Fountain, Inc., 2001–2003. He is currently Esther and Harold E. Edgerton Assistant Professor of electrical engineering at the Massachusetts Institute of Technology, Cambridge. His research interests include source coding theory, sampling, quantization, and information gathering and dispersal in networks.

Dr. Goyal is a member of Phi Beta Kappa, Tau Beta Pi, Sigma Xi, Eta Kappa Nu and SIAM. In 1998, he received the Eliahu Jury Award of the University of California, Berkeley, awarded to a graduate student or recent alumnus for outstanding achievement in systems, communications, control, or signal processing. He was also awarded the 2002 IEEE Signal Processing Society Magazine Award and an NSF CAREER Award. He serves on the IEEE Signal Processing Society's Image and Multiple Dimensional Signal Processing Technical Committee and as a permanent Conference Co-chair of the SPIE Wavelets conference series.

Kannan Ramchandran (S'92–M'93–SM'98–F'05) received the Ph.D. degree in electrical engineering from Columbia University, New York, in 1993.

He has been a Professor in the EECS at the University of California at Berkeley since 1999. From 1993 to 1999, he was on the faculty of the ECE Department at the University of Illinois at Urbana-Champaign. Prior to that, he was a Member of the Technical Staff at AT&T Bell Laboratories, Whippany, NJ, from 1984 to 1990. His current research interests include distributed signal processing architectures and decentralized coding algorithms for sensor and ad hoc wireless networks, multi-user information theory, content delivery for peer-to-peer and wireless networks, media security and information-hiding, and multiresolution statistical image processing and modeling.

Dr. Ramchandran received the Eliahu I. Jury Award in 1993 at Columbia University for the best doctoral thesis in the area of systems, signal processing, and communications, an NSF CAREER award in 1997, the ONR and ARO Young Investigator Awards in 1996 and 1997 respectively, the Henry Magnuski Outstanding Young Scholar Award from the ECE Department at UIUC in 1998, and the Okawa Foundation Prize from the EECS Department at UC Berkeley in 2001. He is the co-recipient of two Best Paper Awards from the IEEE Signal Processing Society (1993, 1997), and of several best paper awards at conferences and workshops in his field. He serves on numerous technical program committees for premier conferences in information theory, communications, and signal and image processing. He was co-chair of the Third IEEE/ACM Symposium on Information Processing for Sensor Networks (IPSN) at Berkeley in 2004. He has been a member of the IEEE Image and Multidimensional Signal Processing Technical Committee and the IEEE Multimedia Signal Processing Technical Committee, and has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING.