

ROBUSTNESS TO CHANNEL VARIATION IN SOURCE CODING FOR TRANSMISSION ACROSS NOISY CHANNELS

Michelle Effros

Dept. of Electrical Engineering, 136-93
California Institute of Technology
Pasadena, CA 91125
effros@caltech.edu

ABSTRACT

We consider the problem of lossy source coding for transmission across an unknown or time-varying noisy channel. The objective is to design an optimal compression system for applications where the unknown channel characteristics are independently estimated at the channel encoder and decoder. Channel estimation reliability is allowed to vary from perfect channel identification to no channel identification. In each case, the goal in system design and operation is to achieve the best possible expected performance with respect to the unknown channel state and the accuracy of the channel estimators. We describe an optimal design technique and an algorithm for achieving optimal expected performance for the entire array of channel estimation accuracies. The resulting system achieves up to 9 dB improvement over the performance on a system designed assuming zero probability of error when used to encode a collection of medical brain scans for transmission across a finite state channel containing two equally probable binary symmetric channels with crossover probabilities .05 and .001.

1. INTRODUCTION

Shannon's separation theorem and its later generalizations prove that for very broad families of sources and channels, independently designed source and channel codes yield the same performance capabilities as jointly designed codes. As a result, the problems of source and channel coding have largely been addressed independently. Unfortunately, the separation theorem is an asymptotic result. Thus while it is true that the optimal performance of a communication system containing independently designed source and channel codes equals the optimal performance of a combined source and channel coder at infinite vector dimensions and block lengths, the same cannot be said for independently designed source and channel coders in the finite dimensional world of practical codes.

As a result, a number of researchers in both the source and channel coding communities have considered the problem of joint source and channel code design for practical applications. In this paper, we concentrate on a subset of the above joint design problem. In particular, we consider

the problem of fixed rate lossy source code design and operation for scenarios where the lossily encoded source symbols will be subjected to noise.

A number of authors have studied the problem of source code design for transmission across noisy channels (e.g., [1] and the references therein). In [1], Farvardin analyzes the optimal design algorithm and associated performance for sending the output of a fixed-rate vector quantizer (VQ) across a noisy channel. Unfortunately, as Farvardin notes, the optimal design and performance are channel dependent. Thus, a source code designed for one channel and operated across another channel will not achieve the optimal performance for the channel in operation. As a result, Farvardin's code design is ill-suited for many of today's communication applications, where channel characteristics are often unknown at design time or even time-varying.

In order to study source code design for channels with unknown or time-varying statistics, we must first develop a reasonable channel model. We here consider a discrete, finite-state channel model. In this model, the channel is considered to pass through some finite number of possible "states" or modes. The channel transition probabilities for any given channel use are a deterministic function of the channel's state at the time of that use. In modeling a channel of unknown characteristics, the finite collection of states and state probabilities is chosen according to some reasonable quantization of the space of possible channels and probability of each of the resulting channel models. A time-varying channel may similarly be modeled by choosing a collection of channels and modeling the random process according to which the channel changes among the states.

In [2], Duman and Salehi consider the problem of source code design for the above described finite-state channel model. In that work, they describe an algorithm for achieving optimal performance on the *best* channel within the finite-state channel's collection of possible channels subject to a constraint on the performance on the *worst* channel in the collection. In this work, we replace the above constrained optimization with a simple unconstrained optimization. The algorithm described in Section 3 minimizes the overall *expected* end-to-end distortion through the system for each possible fixed-rate source code, where the expectation takes into account both the random process describing the channel probabilities and the inaccuracy in the estimation of the channel state at the encoder and decoder.

This material is based upon work supported by NSF Grant No. MIP-9501977

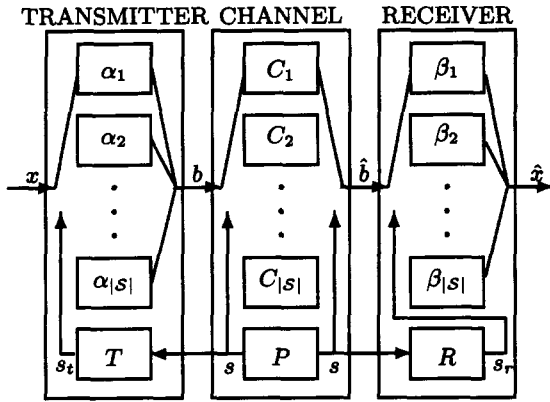


Figure 1: The communication system model.

2. THE PROBLEM

Consider a discrete channel with fixed-rate binary input $b \in \{0, 1\}^n$ and output $\hat{b} \in \{0, 1\}^n$. Let $\mathcal{S} = \{1, 2, \dots, |\mathcal{S}|\}$ be a finite collection of channel state indices. We assume that the channel switching process is a memoryless random process that changes state at most once every channel use. That is, we assume that the channel state will not change in the midst of a single channel transmission and that the next state function is an iid random process with alphabet \mathcal{S} . For any $s \in \mathcal{S}$ and any $b, \hat{b} \in \{0, 1\}^n$, let $P(s)$ be the probability of state s and $C_s(\hat{b}|b)$ be the probability that the channel output equals \hat{b} given a channel input of b and channel state s . Since the channel state is described by a random process and this process is unknown to the source encoder and decoder, we include a state estimator at both the encoder and decoder. The state estimators will likewise be modeled as random processes. For any $s, \hat{s} \in \mathcal{S}$, we define $T(\hat{s}|s)$ and $R(\hat{s}|s)$ to be the probability that the transmitter and receiver respectively estimate the channel state to equal \hat{s} given a true channel state of s .

As in [2], we will allow the encoder and decoder to change as a function of the transmitter and receiver state estimations s_t and s_r . We therefore will consider the design of a collection of encoders $\{\alpha_s : s \in \mathcal{S}\}$ and a collection of decoders $\{\beta_s : s \in \mathcal{S}\}$. For each $s \in \mathcal{S}$, the encoder $\alpha_s : \mathcal{X}^k \rightarrow \{0, 1\}^n$ maps the input space \mathcal{X}^k of k -dimensional data vectors to the output space $\{0, 1\}^n$ of binary n -tuples. Likewise, each decoder $\beta_s : \{0, 1\}^n \rightarrow \hat{\mathcal{X}}^k$ maps its input space of possible binary n -tuples to the output space of possible reproduction values. (Typically, $\hat{\mathcal{X}} \subseteq \mathcal{X}$.) The resulting system is illustrated in figure 2.

Suppose that we are given some probability distribution function q on the space of possible input vectors and a distortion measure $d : \mathcal{X}^k \times \hat{\mathcal{X}}^k \rightarrow \mathbb{R}$. We will use $d(x^k, \hat{x}^k) = \sum_{i=1}^k (x_i - \hat{x}_i)^2$ throughout this paper. Then the expected distortion associated with using the above code is

$$D = \int_{\mathcal{X}^k} \sum_{s=1}^{|\mathcal{S}|} \sum_{s_t=1}^{|\mathcal{S}|} \sum_{s_r=1}^{|\mathcal{S}|} \sum_{\hat{b} \in \{0, 1\}^n} q(x^k) P(s) T(s_t|s) R(s_r|s) C_s(\hat{b}|\alpha_{s_t}(x^k)) d(x^k, \beta_{s_r}(\hat{b})) dx^k.$$

In the following section we address the problem of optimal encoder and decoder design given an arbitrary finite-state channel and pair of estimation error distributions for any fixed rate n .

3. THE ALGORITHM

Consider the problem of code design for the communications system described in the previous section. Notice that at any given time, the transmitter has a single estimate s_t of the current channel state and uses the encoder α_{s_t} associated with the current estimate. The optimal encoder α_{s_t} is thus the encoder that minimizes for each x^k the expected distortion between x^k and its reproduction at the decoder given that the transmitter state estimate is s_t . Thus for each $s_t \in \mathcal{S}$, we describe the optimal encoder $\alpha_{s_t}^*$ as

$$\alpha_{s_t}^*(x^k) = \arg \min_{b \in \{0, 1\}^n} \sum_{s=1}^{|\mathcal{S}|} \sum_{s_r=1}^{|\mathcal{S}|} \sum_{\hat{b} \in \{0, 1\}^n} P(s) T(s_t|s) R(s_r|s) C_s(\hat{b}|b) d(x^k, \beta_{s_r}(\hat{b})).$$

Notice that the optimal encoder is no longer the nearest neighbor encoder of traditional vector quantization. The binary string to which $\alpha_{s_t}^*$ maps a given x^k vector may not be the one with the closest reproduction $\beta_{s_r}(\cdot)$ due to the fact that the binary string may be corrupted in transmission and the channel estimates may be incorrect at either the encoder or the decoder or both. Thus every decoder and every codeword in each of those decoders affects the encoder's optimal encoding rule.

The above optimal encoding rule holds across all possible accuracies for the transmitter's state estimator. If the encoder's state estimation is perfect, then $T(s_t|s) = 1$ if $s_t = s$ and 0 otherwise, and the encoding rules for the $|\mathcal{S}|$ encoders in the collection may be quite different. If $T(s_t|s) = 1/|\mathcal{S}|$ for all $s, s_t \in \mathcal{S}$, then the encoders should be identical for each value of s_t .

Given a collection of optimal encoders and a pair of estimation accuracies, we next consider the optimal collection of decoders for the given system. For any $s_r \in \mathcal{S}$ and $\hat{b} \in \{0, 1\}^n$, the receiver will use codeword $\beta_{s_r}(\hat{b})$ only when its state estimate equals s_r and the channel output equals \hat{b} . Thus the optimal collection of decoders must have the property

$$\beta_{s_r}^*(\hat{b}) = \arg \min_{\hat{x}^k} \int_{\mathcal{X}^k} \sum_{s=1}^{|\mathcal{S}|} \sum_{s_t=1}^{|\mathcal{S}|} q(x^k) P(s) R(s_r|s) T(s_t|s) C_s(\hat{b}|\alpha_{s_t}(x^k)) d(x^k, \hat{x}^k) dx^k.$$

For the case of the squared error distortion measure, the resulting codeword equals the expected value of x^k given that the channel output is \hat{b} and the state estimate at the receiver is s_r . Thus

$$\beta_{s_r}^*(\hat{b}) = C \int_{\mathcal{X}^k} \sum_{s=1}^{|\mathcal{S}|} \sum_{s_t=1}^{|\mathcal{S}|} q(x^k) P(s) R(s_r|s) T(s_t|s) C_s(\hat{b}|\alpha_{s_t}(x^k)) x^k dx^k$$

where C is a normalizing constant equal to

$$C = \left[\int_{\mathcal{X}^k} \sum_{s=1}^{|\mathcal{S}|} \sum_{s_t=1}^{|\mathcal{S}|} q(x^k) P(s) R(s_r|s) T(s_t|s) C_s(\hat{b}|\alpha_{s_t}(x^k)) dx^k \right]^{-1}.$$

Again, the codewords have changed from their traditional VQ values. Each codeword $\beta_s(\hat{b})$ is no longer the centroid of its encoding region $\alpha_s^{-1}(\hat{b})$, but rather a weighted average of input values that may be mapped to the given cell, possibly through channel or estimator error.

Like the optimal encoding rule, the optimal decoding rule holds for estimators ranging from perfect to nonexistent. In the case of perfect estimation at both the encoder and the decoder, the codeword values are weighted sums of the centroids of the encoding regions of the same index s . (The weighting results from the transition probabilities associated with the given channel.) In the case of imperfect estimation at either the transmitter or the receiver, the codeword values become influenced by symbols from all possible encoding regions.

The optimal design algorithm is an iterative descent technique analogous to the generalized Lloyd algorithms. We initialize the system with an arbitrary collection of encoders $\{\alpha_s : s \in \mathcal{S}\}$ and decoders $\{\beta_s : s \in \mathcal{S}\}$. Each iteration proceeds through two steps, enumerated below.

1. *Optimize each encoder* for the given collection of encoders and decoders.
2. *Optimize each decoder* for the given collection of encoders and decoders.

Each of the above steps decreases the expected distortion. Since the expected distortion cannot be negative, the algorithm is guaranteed to converge. Since each of the above steps finds a global minimum, the algorithm as a whole, when run to convergence, yields a local minimum.

4. EXPERIMENTAL RESULTS

We compare the expected performance of the channel optimized VQ at a variety of estimation accuracies to the performance of standard VQ sent across the same channel. We set $|\mathcal{S}| = 2$ and $P(1) = P(2) = 0.5$ in all examples. The channels considered are binary symmetric channels with crossover probabilities 0.05 and 0.001. In each case, the encoder is trained on a collection of 20 medical brain scans and then tested on a collection of 5 scans outside of the training set. We use a vector dimension of 4 for all codebooks.

Figure 2 is a graph of the performance of the channel optimized VQ with encoder and decoder channel estimation ranging from complete knowledge of the channel state at both encoder and decoder to no knowledge of the channel state at either the encoder or the decoder. The performance of the channel optimized VQ with less than perfect knowledge of the channel state at both encoder and decoder always matches (at very low rates) or exceeds the performance of the codes which are not optimized for the channel model. In most cases, the performance of the channel optimized codes with less than perfect knowledge of the channel

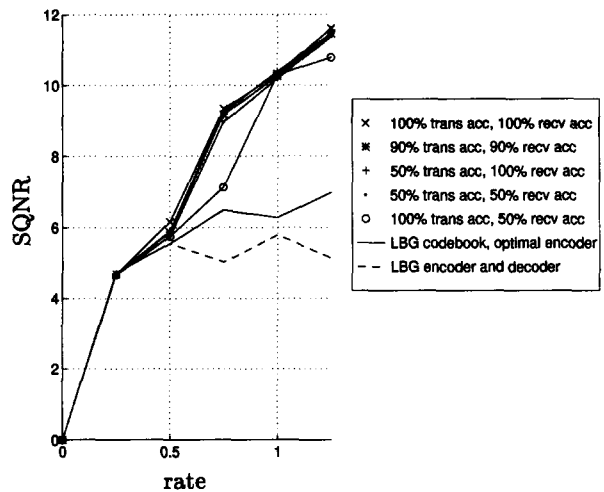


Figure 2: Coding results given channel estimation at the transmitter and receiver with a variety of accuracy levels. (Accuracy is defined as the percentage of the time that the estimator correctly estimates the channel state.)

state is almost identical to the performance of the perfect knowledge code. Compressed images resulting from joint source and channel coding codebooks using perfect estimation, no estimation, and a codebook designed assuming no channel errors (but encoded with the optimum encoder) are shown in Figure 3.

To better understand the above results, we plot in Figure 4 the effect of mismatch between the codebook and channel. In this graph, we compare the rate-distortion performance of a codebook trained for and operated on a channel with probability of error 0.001 with the performance of a codebook trained for a channel with probability of error 0.05 and operated on a channel with probability of error 0.001. Likewise, we compare the rate-distortion performance of a codebook trained for and operated on a channel with probability of error 0.05 with the performance of a codebook trained for a channel with probability of error 0.001 and operated on a channel with probability of error 0.05. We notice that on the given data set, the performance suffers very little when the codebook design is pessimistic (assumes a higher probability of error than necessary) but quite a bit when the codebook design is optimistic. Thus we pay little in rate distortion performance for adding extra error protection to our codebook. As a result, there exists a single (pessimistic) codebook that does well on both channel states.

We further observe in Figure 2 that the codebook with perfect channel knowledge at the transmitter and no channel knowledge at the receiver shows, at two rates, worse performance than that of the code with no knowledge at either transmitter or receiver. This result is clearly a local minimum problem that could be avoided by initializing the codebooks in this case with the codebooks from the code with no knowledge at encoder and decoder.

Finally, Figure 5 shows the results achieved by a simple modification to the design algorithm. The modification, de-

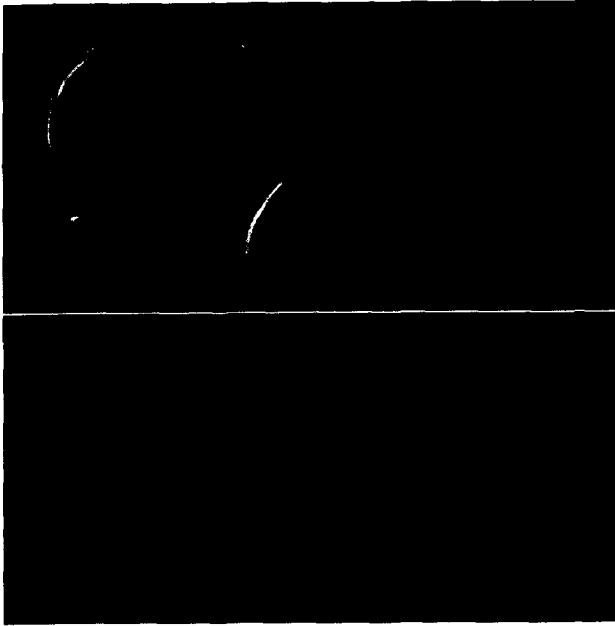


Figure 3: Compressed images at .75 bpp. Top row: Original, LBG codebook with optimal encoder; Bottom row: joint source channel codes with no channel estimation and perfect channel estimation.

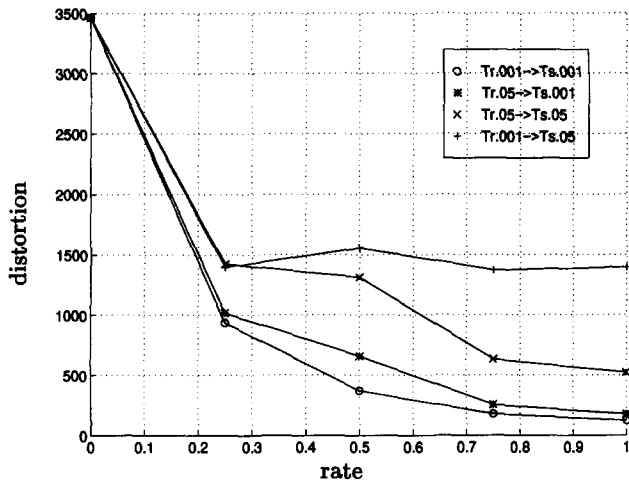


Figure 4: Coding results comparing the performance of a single codebook trained and tested on channels with probability of error 0.05 and 0.001.

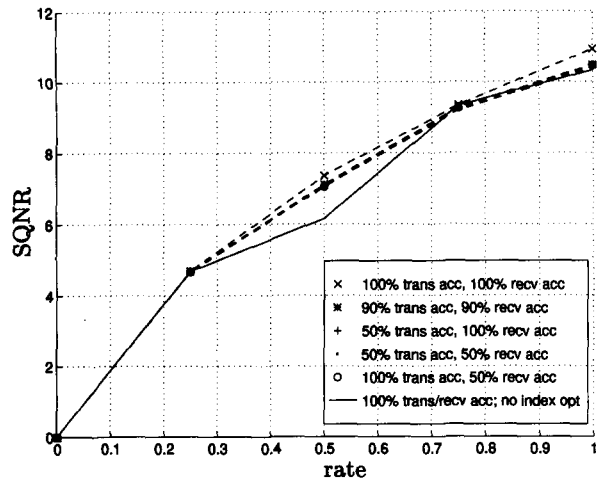


Figure 5: Coding results with index optimization compared to the best performance (100% accuracy at both encoder and decoder) without index optimization.

scribed next, seems to be quite effective in reducing the algorithm's problems with local minima. Notice in the above described algorithm that if we choose the binary representation associated with a given codeword as some function of the codeword's index, then the index choice affects the probability that a given codeword, when transmitted through the channel, will be flipped to each other codeword in the collection. One possible approach towards alleviating our problems with local minima would be to add a third step to the design algorithm in which we attempt to improve the codeword indices (or equivalently their binary representations). The results of such an addition are chronicled in Figure 5. As desired, the design algorithm modification seems to totally alleviate our local minima problem on the given data set. With complete knowledge of the channel state at the transmitter and receiver, the channel optimized VQ achieves up to 9 dB improvement over the standard VQ and 4 dB improvement over the standard VQ codebook with a channel optimized encoder. The joint codes with imperfect channel knowledge at the transmitter and receiver achieve similar gains.

5. REFERENCES

- [1] N. Farvardin. A study of vector quantization for noisy channels. *IEEE Transactions on Information Theory*, IT-36(4):799-809, July 1990.
- [2] T. Duman and M. Salehi. Optimal quantization for finite state channels. In *Proceedings of the IEEE International Symposium on Information Theory*, page 378, Whistler, British Columbia, September 1995. IEEE.