

Design of Optimal Quantizers for Distributed Source Coding

David Rebollo-Monedero, Rui Zhang and Bernd Girod
Information Systems Laboratory, Electrical Eng. Dept.
Stanford University, Stanford, CA 94305
{drebollo,rui,bgirod}@stanford.edu

Abstract

We address the problem of designing optimal quantizers for distributed source coding. The generality of our formulation includes both the symmetric and asymmetric scenarios, together with a number of coding schemes, such as the ideal coding achieving a rate equal to the joint conditional entropy of the quantized sources given the side information. We show the optimality conditions quantizers must satisfy, and generalize the Lloyd algorithm for its design. Experimental results are shown for the Gaussian scalar asymmetric case.

1 Introduction

Consider a network of low-cost remote sensors sending data to a central unit, which may also have access to some side information, for instance, previous data or data provided by local sensors. Suppose that at each remote sensor the data acquired by the other sensors and the side information is not available, but the statistical dependencies are known. We can nevertheless exploit these dependencies to achieve more efficient source coding. If lossy coding is allowed, a convenient coding implementation consists of a quantizer and a lossless coder. In fact, rate-distortion theory, both for non-distributed source coding and source coding with side information at the decoder [1, 2, 3], guarantees that block codes of sufficiently large length are asymptotically optimal, and they can be seen as vector quantizers followed by fixed-length coders.

Clearly, both the quantizers and the lossless coders may benefit from exploiting the statistical dependencies. Practical distributed lossless coding schemes have been proposed (see, e.g. [4, 5]) that are getting close to the Slepian-Wolf bound [6]. As for the design of optimal quantizers, some recent approaches are heuristic [7] or consider only a subset of solutions, such as partitions with connected quantization regions [8]. A more general treatment of the problem of quantizer design is presented in [9], where the Lloyd algorithm [10] is extended to distortion-only optimized quantizers for network source coding. [11] (unpublished) is a further extension incorporating Lagrangian rate-distortion cost functions. This latter work deals with rates defined as expectations of functions of the quantization indices, mainly expectations of linear combinations of actual codeword lengths, where the codebook is designed taking into account the statistical dependencies among sources and side information.

In this paper, we study the design of quantizers for distributed lossless source coding, optimized in terms of distortion and rate. The main contribution is the inclusion of the case in which the rate equals the joint conditional entropy of the quantization indices given the side information, that is, optimal quantization for Slepian-Wolf coding. This work extends the framework for optimal quantizer design for non-distributed sources [12, 13], especially the Lloyd algorithm. The formulation and the solution of the problem studied here were developed independently of [9, 11], and while there are several similarities in the treatment of the distortion, our framework is more general as far as rate measures is concerned. This greater generality turns out to be essential for the practically important cases where the (low) dimensionality of the quantizer is unrelated to the (large) block length of the lossless coder. Specifically, both the case of actual codeword lengths and the case of joint conditional entropy are covered. There are also important differences in the presentation of the overlapping part of the theory, as well as in the implementation of the algorithm. On the other hand, [9, 11] consider a general network, which includes in particular the symmetric and asymmetric settings.

This paper is organized as follows. In Section 2 the problem of quantizer design is formally stated, and illustrated with several examples of coding schemes. A solution extending the Lloyd algorithm is presented in Section 3. Finally, Section 4 provides simulation results for the Gaussian scalar asymmetric case.

2 Formulation of the Problem and Coding Cases

We study the optimal quantizer design for the distributed source coding setting depicted in Fig. 1. We follow the convention of using uppercase letters for random

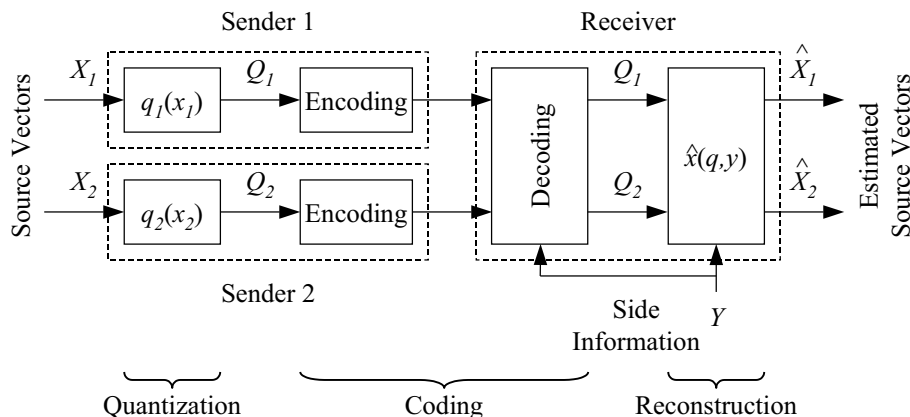


Figure 1: Quantizers for distributed source coding.

vectors, and lowercase letters for particular values these vectors take on. We wish to quantize two vectors produced by two sources, modelled by the random vectors X_1 and X_2 , not necessarily of the same dimension. Define $X = (X_1, X_2)$. Each source vector is available only at its corresponding quantizer. Let the quantization

functions $q_1(x_1)$ and $q_2(x_2)$ map each source vector into a quantization index, Q_1 and Q_2 , grouped as $Q = (Q_1, Q_2)$. A random vector Y , of dimension possibly different from X_1 and X_2 , plays the role of side information available only at the receiver. The side information and both quantization indices are used jointly to estimate the source vectors. Let \hat{X}_1 and \hat{X}_2 represent these estimates, obtained with the reconstruction function $\hat{x}(q, y)$.

The formulation in this work is concerned only with the design of the quantizers and the reconstruction function, not with the coding part. It assumes that the coding of the indices Q_1 and Q_2 with side information Y is lossless, or does not contribute significantly to the distortion. Observe that in general the coding block length is unrelated to the dimension of the vectors X_1 and X_2 .

The function $d(x, \hat{x})$ denotes a distortion measure, and $\mathcal{D} = E[d(X, \hat{X})]$ the expected distortion. Let $r(q, y)$ be a function representing some rate measure. Its role consists in characterizing the coding scheme, as we shall see in this section. The expected rate is then defined as $\mathcal{R} = E[r(Q, Y)]$.

X_1 , X_2 and Y are in general statistically *dependent*. We emphasize that each quantizer only has access to its corresponding source, thus the values of the other source and the side information are unknown. However, the joint statistics of X and Y are assumed to be known, and exploited in the design of the optimal quantizers. We consider the problem of finding the quantizers and reconstruction function that minimize the expected Lagrangian cost $\mathcal{J} = (1 - \lambda) \mathcal{D} + \lambda \mathcal{R}$, where λ is a real number in $[0, 1]$.

We now illustrate the meaning of the rate measure $r(q, y)$ with an example. Suppose that we have the coding setting depicted in Fig. 1, in which each quantization index Q_1 and Q_2 is encoded separately and the side information Y is not available at any of the encoders, but that the statistical dependencies between all of them are known. A consequence of the Slepian-Wolf Theorem [6], specifically of its version for three sources (see, e.g. [3]), is that the set of achievable rates to encode the quantization indices is the same as if both sources were jointly encoded and the side information were available at the encoder. The lowest achievable total rate is $H(Q_1, Q_2|Y)$ in both cases, which is precisely \mathcal{R} when $r(q, y) = -\log p_{Q|Y}(q|y)$. Since both cases share a common definition for the rate measure $r(q, y)$, the design problem studied here is mathematically identical. The second case, however, provides a clearer physical meaning for $r(q, y)$. It represents the ideal codeword length required to code the vector of indices q when the side information takes the value y .

Our framework can be immediately generalized to more than two sources. It can also be easily specialized to the symmetric distributed source coding case, in which there is no side information, to the asymmetric case, where there is only one source, or to the non-distributed case.

Alternative definitions of $r(q, y)$ model different coding settings. Either probabilities or actual codeword lengths can be used in its definition. Some examples are shown in Table 1. In the first three, the Slepian-Wolf bound is assumed to be achieved. These choices do not assume any statistical independence between Q_1 , Q_2 and Y , they merely model different coding schemes.

Suppose that the coding part is carried out jointly on blocks of quantization

$r(q, y)$	\mathcal{R}	Coding
$-\log(p_{Q Y}(q y))$	$H(Q_1, Q_2 Y)$	Distributed coding.
$-\log(p_Q(q))$	$H(Q_1, Q_2)$	Symmetric distributed coding, side information ignored.
$-\log(p_{Q_1 Y}(q_1 y) p_{Q_2 Y}(q_2 y))$	$H(Q_1 Y) + H(Q_2 Y)$	Asymmetric distributed coding, source dependence ignored.
$-\log(p_{Q_1}(q_1) p_{Q_2}(q_2))$	$H(Q_1) + H(Q_2)$	Separate encoding, all dependencies ignored.
$(1 - \mu) l_1(q_1) + \mu l_2(q_2)$	$(1 - \mu) E[l_1(Q_1)] + \mu E[l_2(Q_2)]$	Use of a specific codebook with codeword lengths $l_i(q_i)$. Rates are weighted.
$a p_{Q_1}(q_1) + b p_{Q_2}(q_2) + c p_Q(q y)$	$a H(Q_1) + b H(Q_2) + c H(Q Y)$	Linear combination of previous cases.

Table 1: Some examples of rate measures $r(q, y)$ and their applications.

indices. Then, each of the codeword lengths l_1 and l_2 would depend on several samples of the quantization indices q_1 and q_2 , and $r(q, y)$ could *not* be written simply as $l_1(q_1) + l_2(q_2)$. Grouping source vectors in blocks of the same size as those used in the coding would solve this problem, but it would increase the dimensionality and therefore the complexity of the quantizer design.

Consider now the method for coding a source vector X_1 with side information Y , based on turbo codes, as reported in [5]. The technique works with extremely large fixed-length sequences, and the rate is fixed by the convolutional code used. The probability of decoding error increases steeply as the conditional entropy of the source vector given the side information approaches this fixed rate. This particular coding scheme can be seen as nearly lossless, with a rate close to the conditional entropy. A first model would be a lossless coder with a rate $\mathcal{R} = H(Q_1|Y)$, or perhaps more accurately, $\mathcal{R} = a H(Q_1) + b H(Q_1|Y)$.

3 Optimal Quantizer Design

The functions defined below represent the expected distortion, expected rate and expected Lagrangian cost associated to a particular mapping of a source value into a quantization index:

$$\begin{aligned}
\tilde{d}_1(x_1, q_1) &= E[d((x_1, X_2), \hat{x}((q_1, Q_2), Y)) | X_1 = x_1] \\
\tilde{d}_2(x_2, q_2) &= E[d((X_1, x_2), \hat{x}((Q_1, q_2), Y)) | X_2 = x_2] \\
\tilde{r}_1(x_1, q_1) &= E[r((q_1, Q_2), Y) | X_1 = x_1] \\
\tilde{r}_2(x_2, q_2) &= E[r((Q_1, q_2), Y) | X_2 = x_2] \\
\tilde{j}_1(x_1, q_1) &= (1 - \lambda) \tilde{d}_1(x_1, q_1) + \lambda \tilde{r}_1(x_1, q_1) \\
\tilde{j}_2(x_2, q_2) &= (1 - \lambda) \tilde{d}_2(x_2, q_2) + \lambda \tilde{r}_2(x_2, q_2)
\end{aligned} \tag{1}$$

For instance, $\tilde{d}_1(x_1, q_1)$ is an estimate of the distortion when the source X_1 emits the value x_1 , if the quantization index corresponding to this value, $q(x_1)$, is chosen to be q_1 , for a particular distortion measure and a particular reconstruction function.

In the entropy-constrained, non-distributed setting, the reconstruction functions become reconstruction levels $\hat{x}_1(q_1)$. For a quadratic distortion measure, $\tilde{d}_1(x_1, q_1)$ is simply the paraboloid $\|x_1 - \hat{x}_1(q_1)\|^2$, and $\tilde{r}_1(x_1, q_1) = -\log p_{Q_1}(q_1)$ is independent from x_1 . In this case, the index q_1 minimizing the cost $\tilde{j}_1(x_1, q_1)$ is chosen. In this section we shall discover how similar the distributed quantizer design is to the non-distributed one. Observe also the similarity of our definitions with the modified distortion measures used in quantization of noisy sources [14, 13], where the distortion between a noise-corrupted observation V of an unseen original U and its reconstruction \hat{U} is defined as $E[d(U, \hat{U})|V]$.

A fundamental property of the distortion, rate and Lagrangian cost functions (1) is that their expectation is precisely the expected distortion, rate and cost, respectively:

$$\begin{aligned}\mathcal{D} &= E[\tilde{d}_1(X_1, Q_1)] = E[\tilde{d}_2(X_2, Q_2)] \\ \mathcal{R} &= E[\tilde{r}_1(X_1, Q_1)] = E[\tilde{r}_2(X_2, Q_2)] \\ \mathcal{J} &= E[\tilde{j}_1(X_1, Q_1)] = E[\tilde{j}_2(X_2, Q_2)]\end{aligned}\tag{2}$$

This property plays a key role in the understanding of the necessary optimality conditions for the quantizer, the reconstruction function and the rate measure, which we shall now show.

For each $i \in \{1, 2\}$, consider all possible quantization functions $q_i(x_i)$, leaving fixed the other quantization function, the reconstruction function $\hat{x}(q, y)$ and the rate measure $r(q, y)$. Then,

$$q_i^*(x_i) = \arg \min_{q_i} \tilde{j}_i(x_i, q_i)\tag{3}$$

defines an optimal quantization function, since it minimizes \mathcal{J} over all choices of the quantization function $q_i(x_i)$. An essential difference with respect to the non-distributed quantizer is that even if the distortion measure is quadratic, it turns out that the estimated distortion functions are not paraboloids in general, and the quantization regions might be disconnected, as it will be shown in Section 4. Consider for instance the fixed-rate asymmetric case in which X_2 is irrelevant, with a fixed number of quantization indices for X_1 . Suppose that X_1 is of dimension one, and Y is its noisy version, taking values very close to X_1 with very high probability. Then, if two quantization intervals far away from each other are assigned to a common quantization index, the reconstruction process should be able to determine—with high probability—which interval X_1 belongs to. For a fixed number of quantization indices, assigning multiple intervals to a common index allows the quantization function to work with narrower intervals, thus reducing the distortion.

For some particular quantization functions $q_1(x_1)$ and $q_2(x_2)$,

$$\hat{x}^*(q, y) = \arg \min_{\hat{x}} E[d(X, \hat{x})|Q = q, Y = y]\tag{4}$$

is an optimal reconstruction function, since it minimizes \mathcal{D} over all choices of the reconstruction function $\hat{x}(q, y)$. Furthermore, if $d(x, \hat{x}) = \|x - \hat{x}\|^2$, then

$$\hat{x}^*(q, y) = E[X|Q = q, Y = y].\tag{5}$$

Let $p_{Q|Y}(q|y)$ denote the conditional PMF of Q given Y , and let $p_{S|Y}(s|y)$ denote the conditional PMF of a certain random vector $S = (S_1, S_2)$ given Y . Here S is in general different from Q , but it also assumes values in the set of possible quantization indices. Fixing both quantization functions $q_1(x_1)$ and $q_2(x_2)$,

$$r^*(q, y) = -\log p_{Q|Y}(q|y) \quad (6)$$

is an optimal rate measure, since it minimizes \mathcal{R} over all choices of rate measures $r(q, y)$ of the form $-\log p_{S|Y}(q|y)$ (evaluated at q , not at s). Furthermore, $\mathcal{R}^* = H(Q|Y)$. There are analogous results for each of the alternative definitions of $r(q, y)$ in terms of logarithms of probabilities shown in Table 1.

The following result gives an important simplification for the estimated rate functions when l_1 and l_2 are actual codeword lengths, or ideal codeword lengths as in

$$r(q, y) = -\log(p_{Q_1}(q_1)p_{Q_2}(q_2)), \quad (7)$$

where $l_i(q_i) = -\log p_{Q_i}(q_i)$ and $\mathcal{R} = H(Q_1) + H(Q_2)$. Suppose that the rate measure is of the form $r(q, y) = l_1(q_1) + l_2(q_2)$ for some functions $l_1(q_1)$ and $l_2(q_2)$. Then for each $i \in \{1, 2\}$ the estimated rate functions can be redefined as

$$\tilde{r}_i(x_i, q_i) = \tilde{r}_i(q_i) = l_i(q_i) \quad (8)$$

without affecting the resulting optimal quantization functions as given in Equation 3.

Having seen the optimality conditions, we are now ready to extend the Lloyd algorithm for the special case of distributed source coding:

1. Choose some initial quantization functions $q_i^{(1)}(x_i)$. Set the iteration counter k to 1.
2. Find the optimal reconstruction function $\hat{x}^{(k)}(q, y)$ for the quantization functions $q_i^{(k)}(x_i)$.
3. Find the optimal rate measure $r^{(k)}(q, y) = -\log p_{Q^{(k)}|Y}(q|y)$ for the quantization indices $Q_i^{(k)} = q_i^{(k)}(x_i)$. Alternative definitions for $r(q, y)$ can also be used, as the ones in Table 1.
4. Compute the expected cost $\mathcal{J}^{(k)}$ associated to the current quantization, rate and reconstruction functions. Depending on its value with respect to the previous one, continue or stop.
5. Find the next optimal quantization functions $q_i^{(k+1)}(x_i)$ for the current reconstruction and rate measure. Increase counter k and go back to 2.

The algorithm defined above satisfies the following properties:

1. The cost is non-increasing, that is, $\forall k \geq 1 \quad \mathcal{J}^{(k+1)} \leq \mathcal{J}^{(k)}$, and it converges to a limit.
2. Any quantizer satisfying the optimality conditions, in particular an optimal quantizer, without ambiguity in any of the minimizations involved, is a fixed point of the algorithm.

3. At the step before finding the new quantization functions the expected rate is the conditional entropy, that is, $\mathcal{R}^{(k)} = H(Q^{(k)}|Y)$.

Note that these properties imply neither that the value of the cost \mathcal{J} the algorithm converges to is a minimum, nor that this cost is a global minimum.

So far we have assumed that the joint statistics of the random vectors X_1 , X_2 and Y were known. Instead, a training set of samples $\{(X_1, X_2, Y)_n\}$ can be provided. If the general form of the PDF or PMF is also unknown, non-parametric statistical techniques (see [15] for an introduction) can be applied to estimate the conditional expectations involved in the generalized Lloyd algorithm, especially if any of the random vectors X_1 , X_2 or Y is continuous. Even if the joint statistics are available, since the conditional expectations themselves can be seen as estimates, nonparametric regression techniques, based or not on kernel functions, such as local mean, local linear regression, smoothing splines, wavelets or projection pursuit regression, are particularly useful. These techniques may be combined with, or already include, dimensionality reduction methods of the variables we condition on, such as principal component analysis or vector quantization.

4 Experimental Results for the Gaussian Scalar Asymmetric Case

In this section, the previous analysis is illustrated with simulation results for the case of scalar quantization of a Gaussian random variable for asymmetric source coding. Let $X \sim \mathcal{N}(0, \sigma_X^2 = 1)$ and $Z \sim \mathcal{N}(0, \sigma_Z^2)$ represent the source and independent noise respectively, and let the side information be $Y = X + Z$. The notation used here is equivalent to the previous one except for the subindex of X_1 and Q_1 , which is omitted. X_2 and Q_2 are irrelevant, for instance constants. Define the input signal to noise ratio as $\text{SNR}_{\text{IN}} = \sigma_X^2 / \sigma_Z^2$, and the output signal to noise ratio as $\text{SNR}_{\text{OUT}} = \sigma_X^2 / \mathcal{D}$. Whenever the number of bins is referred to, only those in $[-6\sigma_X, 6\sigma_X]$ are counted.

Two examples of quantizers obtained with the algorithm are represented in Fig. 2. Observe that in both cases there are disconnected quantization regions, as we mentioned in Section 3, and that the entropy-constrained quantizer is almost uniform. Fig. 3 shows an example of reconstruction function and estimated cost function. The

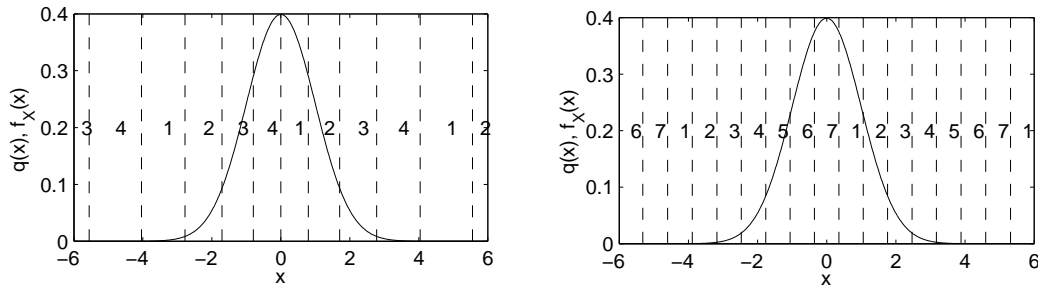


Figure 2: Optimized quantizer for $\lambda = 0$, 4 quantization indices, and $\text{SNR}_{\text{IN}} = 5$ dB (left plot), and $\lambda = .037$, $\mathcal{R} = H(Q|Y)$ and $\text{SNR}_{\text{IN}} = 10$ dB (right plot).

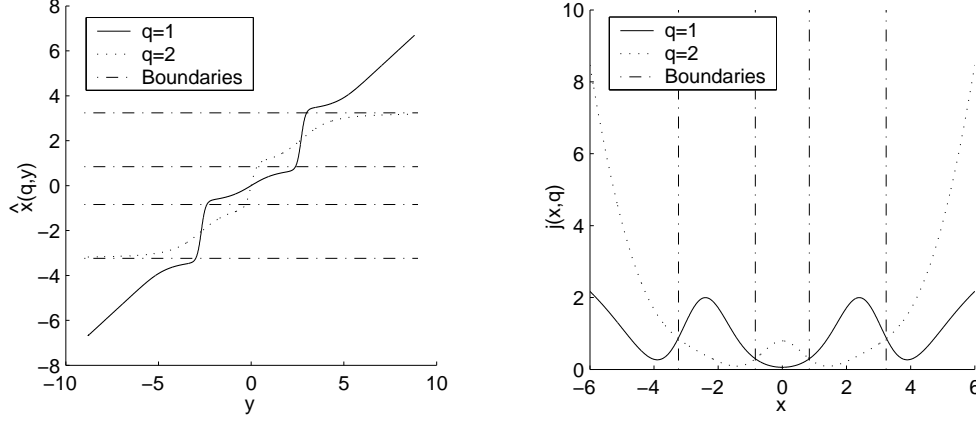


Figure 3: Example of reconstruction functions (left plot) and cost functions (right plot) for $\lambda = 0$, 2 quantization indices and $\text{SNR}_{\text{IN}} = 5$ dB. The quantization boundaries are also shown.

experiments carried out indicate that if the initial quantization function has a number of intervals and quantization indices larger than the optimal, the algorithm converges to the optimal quantizer or to one with nearly optimal cost.

Next, we compare the performance of the following quantization schemes with the Wyner-Ziv bound [2]:

- Ideal quantization according to the **Wyner-Ziv** bound.
- Scalar quantization of the **conditional** random variable $X|Y$. The side information is available at the quantizer.
- **Asymmetric** scalar quantization according to the framework presented here.
- The scalar quantizer is designed as if the source were **independent** from the side information, that is, the optimal non-distributed quantizer for X is used. The reconstruction still takes into account the side information.
- The **side information** is **ignored** both in the design of the scalar quantization and in the reconstruction.
- A **uniform** quantizer is obtained from the optimal, asymmetric quantizer by averaging the interval width according to the PDF of X , and using the same number of indices. The reconstruction takes into account the side information.

Fig. 4 shows the variation in the distortion with the correlation between the source input and the side information, when only the distortion is optimized ($\lambda = 0$) and 4 quantization indices are used (left plot), along with the number of intervals per quantization index. In Fig. 5, distortion-rate curves for minimum Lagrangian cost scalar quantizers are plotted. The rate \mathcal{R} has been set to $H(Q)$ (left plot) and $H(Q|Y)$ (right plot), and the input SNR to 5 dB. Both in the distortion optimized case and the case in which the rate is the unconditional entropy, the quantizer obtained with our generalization of the Lloyd algorithm yields a distortion significantly lower than the quantizer designed using the non-distributed Lloyd algorithm, despite the fact that

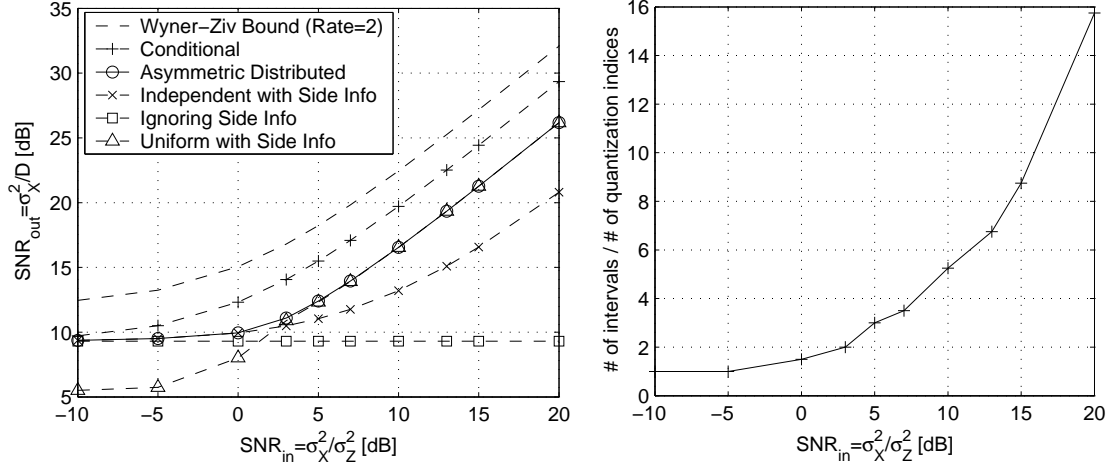


Figure 4: Comparison of distortion optimized quantization schemes using 4 quantization indices ($\lambda = 0$) (left plot), and number of intervals per quantization index vs. input SNR.

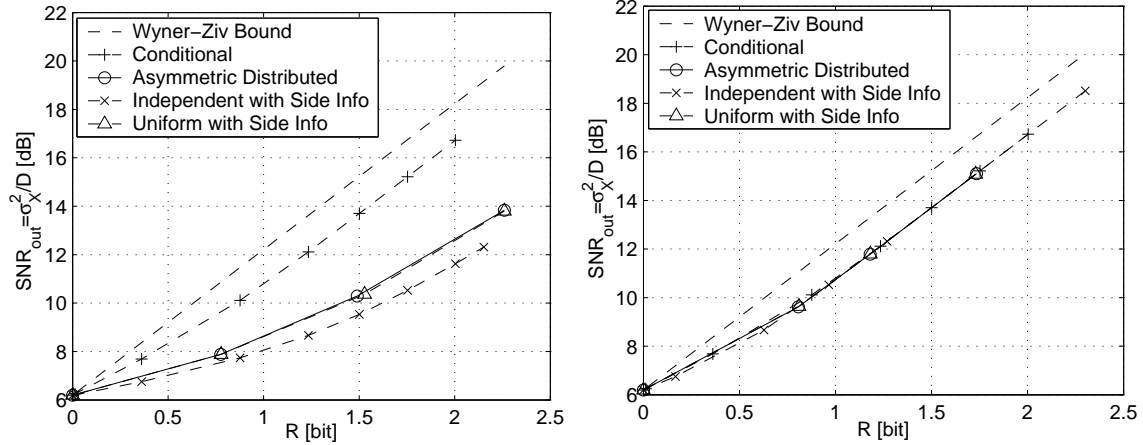


Figure 5: Comparison of quantization schemes with $\mathcal{R} = H(Q)$ (left plot) and $\mathcal{R} = H(Q|Y)$ (right plot), for $\text{SNR}_{\text{IN}} = 5$ dB.

both reconstruction functions use the side information. In the conditional-entropy-constrained case, however, there is no room for improvement, since having the side information available at the quantizer or ignoring it for its design produce almost the same distortion, as long as the coding is ideally efficient. In all cases, when the number of intervals is large, the uniform version of the optimal quantizer performs almost as well. We mentioned in Section 3 that the optimized quantizers might map several bins into a common quantization index. In the distortion optimized case, the number of intervals per quantization index grows with the input SNR (see Fig. 4).

5 Conclusions

A framework for rate-distortion optimized quantizer design for the distributed source coding setting in Fig. 1 has been presented. Along with a distortion measure, a rate

measure has been introduced to model a variety of coding scenarios, including the case of joint conditional entropy of the quantization indices of the sources and the side information (Slepian-Wolf coding), and the case of actual codeword lengths.

The conditions an optimal quantizer must satisfy have been established. The Lloyd algorithm to design a locally optimal quantizer has been generalized, and compared to other schemes in the scalar Gaussian case, showing its better performance.

Acknowledgment

The authors gratefully acknowledge F. Guo for his collaboration in the early development of this theory, and Prof. R.M. Gray for his helpful comments.

References

- [1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pp. 142–163, 1959.
- [2] A. D. Wyner, J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, Jan. 1976.
- [3] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 1991.
- [4] J. García-Frías and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Communication Letters*, vol. 5, no. 10, pp. 417–419, Oct. 2001.
- [5] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conf. (DCC)*, Snowbird, UT, pp. 252–261, Apr. 2002.
- [6] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, Jul. 1973.
- [7] J. Kusuma, L. Doherty and K. Ramchandran, "Distributed compression for sensor networks," in *Proc. International Conf. Image Processing (ICIP)*, Thessaloniki, Greece, vol. 1, pp. 82–85, Oct. 2001.
- [8] D. Muresan and M. Effros, "Quantization as histogram segmentation: globally optimal scalar quantizer design in network systems," in *Proc. IEEE Data Compression Conf. (DCC)*, pp. 302–311, 2002.
- [9] M. Fleming and M. Effros, "Network vector quantization," in *Proc. IEEE Data Compression Conf. (DCC)*, Snowbird, UT, pp. 13–22, Mar. 2001.
- [10] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, Mar. 1982.
- [11] M. Fleming, Q. Zhao and M. Effros, "Network vector quantization," to appear in *IEEE Trans. Inform. Theory*.
- [12] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, Kluwer Academic Publishers, 1992.
- [13] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.
- [14] R. L. Dobrushin and B. S. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inform. Theory*, vol. IT-8, pp. S293–S304, 1962.
- [15] A. W. Bowman and A. Azzalini, *Applied smoothing techniques for data analysis*, Oxford University Press, 1997.