Quantizer Design for Distributed Estimation with Communication Constraints and Unknown Observation Statistics

Vasileios Megalooikonomou, Yaacov Yesha

Abstract— **We consider the problem of quantizer design in a distributed estimation system with communication constraints in the case where only a training sequence is available. Our approach is based on a generalization of regression trees. The lookahead method that we also propose improves significantly the performance. The final system performs similarly to the one that assumes known statistics.**

Keywords— **Quantization, Distributed estimation, Nonlinear estimation, Trees.**

I. Introduction

EXAMPLES of distributed estimation systems can be found in radar and satellite-based remote-sensing systems (LANDSAT) to sonar and seismology. In these systems, an estimation has to be made at the fusion center using data observed at the sensors. Transmitted data has to be compressed due to capacity constraints. We consider the case of two sensors since the generalization to an arbitrary number of sensors is straightforward [9], [10]. The observations of the two sensors are random vectors X_1^q, X_2^r, that are related to an unobservable continuous quantity θ that the fusion center needs to estimate. The sensors do not communicate with each other, and there is no feedback from the fusion center to the sensors. The communication channels are assumed to be error free but with capacity constraints R_1, R_2. Sensor k uses its vector quantizer Q_k to map its observation X_k^b into a fixed length codeword $\hat{X}_k^b$ that is transmitted to the fusion center [7]. The fusion center estimates the parameter θ based on $\hat{X}_k^b$. Many researchers have studied the problem of quantization for distributed estimation in the case where the joint distribution $p(x_1, x_2, \theta)$, is known [13], [9], [10], [4], [3], [1], [6]. Here, we consider a more realistic model where the observation statistics are unknown.

II. Background

Let h be the function of the fusion center that gives the estimate $\tilde{\theta}$ of θ for given quantizers Q_1 and Q_2, i.e., $\tilde{\theta} = h(Q_1(X_1^q), Q_2(X_2^r))$. For the mean-squared error (mse) distortion function, the objective is to find Q_1, Q_2, and h such that the estimation error $Error = E\{[\theta - h(Q_1(X_1^q), Q_2(X_2^r))]^2\}$ is minimized. We follow the concept of *Cooperative Design-Separate Encoding* (see [13]). In the case where the joint distribution $p(x_1, x_2, \theta)$ is known and continuous, necessary conditions for optimal Q_1, Q_2, and h for the mse distortion function are given by Lam and Reibman [10]. In order to find the solution, the Cyclic Generalized Lloyd's Algorithm (CGLA) proposed by Longo, *et al.* [13] in the framework of decentralized hypothesis testing under capacity constraints and for a known joint distribution is used [13], [9], [10], [6]. The CGLA is a variation of the Generalized Lloyd Algorithm (GLA) [11], [12], [5]. The solution to the equations provided by Lam and Reibman is possible because the joint distribution is assumed to be known. However, in the case where only a training sequence is available, the previous method cannot be used. The method that we propose is based on tree–structured predictors called BFOS regression trees [2] [1]. A regression tree has d-dimensional rectangles as nodes. Its root is the entire d-dimensional Euclidian space. The immediate successors of a non-leaf node (rectangle) are the two rectangles obtained by splitting that rectangle by a hyperplane determined by fixing a coordinate $X_k[l]$ to a constant. The term *split* refers to a simultaneous choice of the particular coordinate, and the constant. The leaves are d-dimensional rectangles that partition the space. A predicted value is associated with each leaf. Growing a tree involves optimally spliting the node for which the greatest reduction of the prediction error is noticed.

III. The methods for quantizer and fusion center design

Let $\{(X_1^q, X_2^r)^{(t)}, \theta^{(t)}; t = 1, \ldots, M\}$ be the training set, $\mathcal{T}$, of size M that represents the statistics of the source where t is the index of the training samples. We first build two regression trees, T_1 and T_2, one for each sensor, with a number of leaves m_1 and m_2 respectively. Then we achieve the desired rate by labeling the rectangles with a number of labels (codewords) $n_k = 2^{R_k}$, $n_k \leq m_k$ combining them into the required number of partition regions for each quantizer. Hence, when labels are used, the quantizer partition regions are unions of these

V. Megalooikonomou is with the Department of Computer Science, Dartmouth College, Hanover, NH 03755. E-mail:vasilis@cs.dartmouth.edu . This work was performed while the first author was with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250.

Y. Yesha is with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250. E-mail: yayesha@cs.umbc.edu.

[1] Basic routines for regression trees that we used in the development of our system were found in StatLib (Statistics Department, Carnegie Mellon University), a system for distributing statistical software.

rectangles, otherwise, they are the rectangles themselves. The labeling method produces quantizers with either connected or disconnected partition regions. For scalar observations they are called breakpoint and non-breakpoint quantizers respectively. Let $P_{Q_1} = \{U_i; i = 1, \ldots, N\}$ and $P_{Q_2} = \{V_j; j = 1, \ldots, L\}$, be the regions for X_1^q and X_2^r, respectively. Let $p(x_l)$ be the region to which x_l belongs. The optimal fusion center function h for each pair of codewords i, j is:

$$h(i,j) = \frac{1}{|\mathcal{R}_{i,j}|} \sum_{t:\left(X_1^q, X_2^r\right)^{(t)} \in \mathcal{R}_{i,j}} \theta^{(t)} \tag{1}$$

where $\mathcal{R}_{i,j} = \{(X_1^q, X_2^r)^{(t)} : X_1^{q,t} \in U_i, X_2^{r,t} \in V_j\}$ is a subset of the training set. The estimation error is given by

$$Error = \frac{1}{M} \sum_{t:\left(X_1^q, X_2^r\right)^{(t)} \in \mathcal{T}} \left(\theta^{(t)} - h(p(X_1^{q,t}, p(X_2^{r,t})\right)^2. \tag{2}$$

We now describe the steps of our approach in more detail.

A. Building the regression trees: We build the tree for X_1^q to m_1 leaves given no information about X_2^r and then the tree for X_2^r to m_2 leaves taking into account the already built tree for X_1^q. This procedure involves pruning too. Finding the best split points for one tree considers the existing rectangles of the other selecting each time the split that decreases most the error in the estimate of θ given by Eq. 2. We then improve the trees by first building the tree for X_1^q given the pruned tree for X_2^r and then do the following: 1. Continue to grow from the beginning each tree and prune it given the pruned tree for the other sensor until the error gets larger, 2. Iteratively prune both trees, starting always from the original trees of the previous step, until the fractional drop of the error falls below a given threshold. The first step does not guarantee improvement but the second step does since it always starts from the same original trees of the previous step. However, the first step, even when we stop the iterations when the error gets larger, generally improves more than the second. The fusion center table h is updated between the iterations. When we prune one tree given the other, the risk of every node [2] in the first tree is calculated based on Eq. 2 before the pruning, taking into account the other tree. For pruning we use the *Recursive Optimal Pruning Algorithm* (ROPA) [8]. The purpose of pruning in the fixed rate case is to get a subtree with a given number of leaves having an estimation error as small as possible.

B. Labeling the rectangles: This is a way of grouping the rectangles into partition regions that may be either connected or not connected. In order to label the rectangles that correspond to the leaves of the both trees we propose the s-CGLA (set CGLA), that is related to CGLA. The s-CGLA considers together groups of training samples. Starting with an initial labeling of the rectangles and a fusion center table h for these labels, it iteratively improves the labeling of every quantizer considering the effect of every possible label for every rectangle. It updates h after one complete pass through a quantizer. We also introduce a variation of s-CGLA, the lh-s-CGLA (lookahead s-CGLA) which changes the fusion center temporarily whenever there is a decision that has to be made, in order to calculate the effect of every possible change. Moreover, it also updates h immediately after changing the label of a rectangle of the quantizer so the table h is kept updated all the time. For the initialization of labels we propose two methods: 1. Initialize them randomly (i.e., with probability $1/n_k$ use a label in $0 \ldots (n_k - 1)$), 2. Initialize them by first pruning even more the final pruned subtrees to a number of labels n_1 and n_2 and then assigning the same label to all the rectangles of the previous trees that correspond to one rectangle of the new pruned trees. In the second method the resulting initial quantizers have connected partition regions.

C: Improving the labeled trees: We iteratively build each tree from the beginning, including in this procedure the labels that have been assigned to the rectangles of the other, until the fractional drop of the error falls below a given threshold. The s-CGLA is used to improve the initial labels assigned to each tree.

Outline of our method: We first build the regression trees for the two quantizers. Then we label their rectangles. We repeat the process of growing, pruning, and labeling of one tree given the other labeled tree. The final rectangles are labeled using first the s-CGLA and then the lh-s-CGLA.

IV. Simulation results

We consider the case where the observations are scalar quantities of the form: $x_k = \theta + n_k,\ k = 1, 2,$ where the noises n_k are Gaussian distributed with correlation coefficient ρ and marginal distributions $N(0, \sigma_n^2)$. The parameter θ follows a normal distribution $N(0, 1)$ and is independent of the noises. The training set $\mathcal{T}$ and the test set $\mathcal{T}'$ consist of $20,000$ samples. The error threshold is set to 0.005. The optimal centralized estimator for this case is linear.

Table I shows similar performance for the Lam-Reibman quantizers [9] and for our quantizers with breakpoint initialization of labels. Their quantizers have been restricted to be breakpoint although our quantizers are mainly non-breakpoint. However, they assume a known joint distribution while we do not make such an assumption. Table II presents the improving effect on the estimation error of the various design steps. In *cases 1* and *2* we build both trees to 4 and 32 leaves, respectively, and label the rectangles using s-CGLA. In *case 3* we proceed as in case 2, and we build, prune, and label one tree given the other labeled tree, applying s-CGLA again at the end. In *case 4* we improve over case 3 by applying lh-s-CGLA at the end. Building the trees involves the method that iteratively improves the pruned subtrees. Figure 1 demonstrates that our system is in general superior to the decentralized system that uses the Lloyd-Max quantizers for each observation. This is due to cooperative design. It is

more pronounced when ρ is higher and/or σ_n^2 is low. We also compare the performance of our system to that of the centralized estimator plus Lloyd-Max quantizer which is the optimal centralized system in this case [2]. Figure 2 shows that the improvement in the estimation error decreases for higher values of σ_n^2 and ρ and as the total number of labels increases. For comparison we also present the optimal centralized mse. Finally, we observed that at least one of the quantizers develops disconnected partition regions, i.e., becomes non-breakpoint, as the number of leaves and/or ρ increases.

V. Conclusions

Although we do not make any assumption about the distribution, the performance of our system is similar to that of the Lam-Reibman quantizers that base their design on known distribution. Also, building, pruning, and labeling of one tree given the other labeled tree, and applying the lh-s-CGLA as the last step of the labeling procedure reduce considerably the estimation error (see Table II).

References

[1] E. Ayanoglu, "On Optimal Quantization of Noisy Sources", *IEEE Transactions on Information Theory*, vol. IT-36, no. 6, pp. 1450–1452, Nov. 1990.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth Inc., 1984.

[3] Y. Ephraim and R. M. Gray, "A Unified Approach for Encoding Clean and Noisy Sources by Means of Waveform and Autoregressive Model Vector Quantization", *IEEE Transactions on Information Theory*, vol. IT-34, no. 4, pp. 826–834, Jul. 1988.

[4] T. J. Flynn and R. M. Gray, "Encoding of Correlated Observations", *IEEE Transactions on Information Theory*, vol. IT-33, no. 6, pp. 773–787, Nov. 1987.

[5] R. M. Gray, "Vector Quantization", *IEEE ASSP Magazine*, vol. 1, pp. 4–29, Apr. 1984.

[6] J. A. Gubner, "Distributed Estimation and Quantization", *IEEE Transactions on Information Theory*, vol. IT-39, no. 4, pp. 1456–1459, Jul. 1993.

[7] H. Abut, editor, Vector quantization, in *IEEE Reprint Collection*. Piscataway, NJ: IEEE Press, 1990.

[8] S.-Z. Kiang, R. L. Baker, G. J. Sullivan, and C.-Y. Chiu, "Recursive Optimal Pruning with Applications to Tree Structured Vector Quantizers", *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 162–169, Apr. 1992.

[9] W.-M. Lam and A. R. Reibman, "Quantizer Design for Decentralized Estimation Systems with Communications Constraints", in *Proceedings of the 23rd Annual Conference on Information Sciences and Systems*, Baltimore, MD, Mar. 1989, pp. 489–494.

[10] W.-M. Lam and A. R. Reibman, "Design of Quantizers for Decentralized Estimation Systems", *IEEE Transactions on Communications*, vol. 41, no. 11, pp. 1602–1605, Nov. 1993.

[11] Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design", *IEEE Transactions on Communications*, vol. COM-28, no. 1, pp. 84–95, Jan. 1980.

[12] S. P. Lloyd, "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, vol. IT-28, no. 2, pp. 129–136, Mar. 1982.

[13] M. Longo, T. D. Lookabaugh, and R. M. Gray, "Quantization for Decentralized Hypothesis Testing under Communication Constraints", *IEEE Transactions on Information Theory*, vol. IT-36, no. 2, pp. 241–255, Mar. 1990.

[14] J. K. Wolf and J. Ziv, "Transmission of Noisy Information to a Noisy Receiver with Minimum Distortion", *IEEE Transactions on Information Theory*, vol. IT-16, no. 4, pp. 406–411, Jul. 1970.

[2] It is known [14] that the optimal mean-squared error quantization process is equivalent to first optimally estimating in the mean-squared error sense and then optimally encoding the estimated value with the mean-squared error distortion measure.

TABLE I

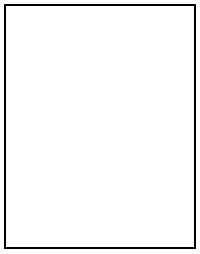
Performance of Lam-Reibman quantizers (a) vs ours (b) for $\sigma_n^2 = 1$.

ρ	0	0.5	0.85
(a)	0.38840	0.47153	0.51734
(b)	0.39599	0.47895	0.51274

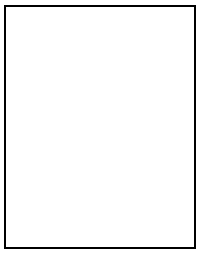
TABLE II

Effect of the various steps in the design (breakpoint initialization, $\rho = 0.85$).

σ_n^2	bp_init, $\rho = 0.85$, (4,4) labels			
	case 1	case 2	case 3	case 4
0.001	0.05515	0.03586	0.03042	0.01385
0.005	0.06024	0.02966	0.02785	0.02325
0.010	0.06318	0.04299	0.04239	0.02369
0.050	0.09134	0.09245	0.07354	0.06107
0.100	0.13937	0.12477	0.12038	0.11105
0.150	0.17504	0.18159	0.15581	0.14499
0.200	0.20962	0.22948	0.20084	0.18366
0.300	0.26941	0.27886	0.25457	0.24643
0.400	0.32167	0.32025	0.31978	0.30149
0.500	0.36532	0.38189	0.36553	0.35600
0.600	0.40913	0.41660	0.40960	0.39220
0.700	0.44504	0.44615	0.44631	0.42747
0.800	0.47186	0.47777	0.46812	0.45892
0.900	0.49771	0.50105	0.49925	0.49197
1.000	0.52223	0.52971	0.52504	0.51274



Vasileios Megalooikonomou was born in Athens, Greece, in 1969. He received the B.S. degree in computer engineering and informatics from the University of Patras, Greece, in 1991 and the M.S. and Ph.D. degrees in computer science from the University of Maryland, Baltimore County, USA, in 1995 and 1997, respectively. From 1997 to 1999 he was a Faculty Research Associate at the Neuroimaging Laboratory, Johns Hopkins University, Baltimore, USA. He is now a Visiting Assistant Professor in the Computer Science Department at Dartmouth College. His interests include data compression, data mining, and multimedia database systems. He is a member of the IEEE Computer Society and a member of the ACM. He was a member of the program committees of the World Multiconference on Systemics, Cybernetics and Informatics (SCI) and the International Conference on Information Systems Analysis and Synthesis (ISAS) 1998, 1999. He has served as a reviewer for several conferences and for the journal Pattern Recognition.



Yaacov Yesha was born in Israel in 1951. He is a Professor at the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County. He received the Ph.D. degree in computer science in 1979, and the M.Sc. degree in computer science in 1974, both from the Weizmann Institute of Science, Rehovot, Israel. In 1973 he received the B.Sc. degree in chemistry from Tel-Aviv University. His interests include scheduling on parallel architectures,

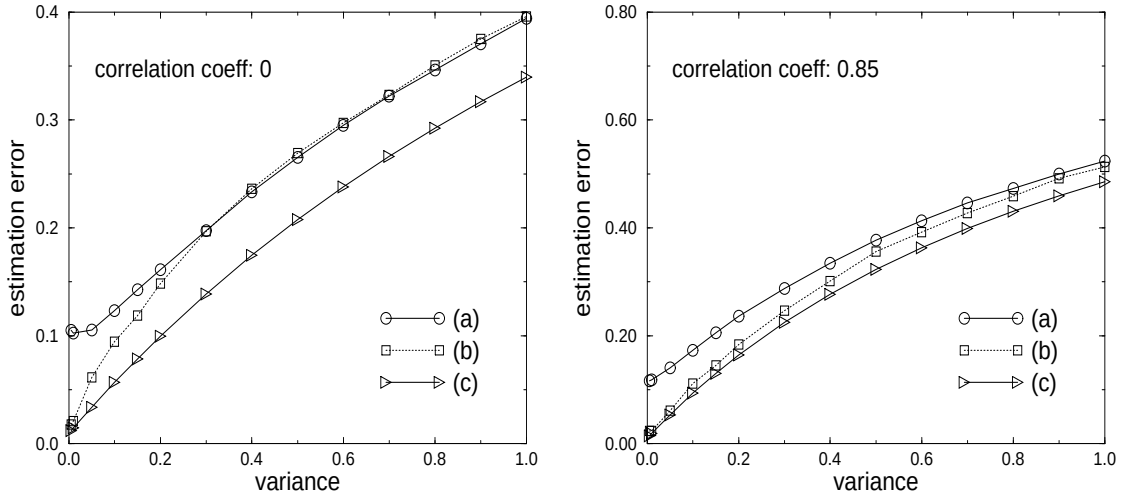


Fig. 1. Performance comparison of (a) Lloyd-Max quantizers for each observation (4 partition regions / quantizer), (b) our quantizers with breakpoint initialization of labels (32 leaves, 4 labels / quantizer), and (c) the centralized estimator plus Lloyd-Max quantizer (16 partition regions), for $\rho = 0$ and $\rho = 0.85$.

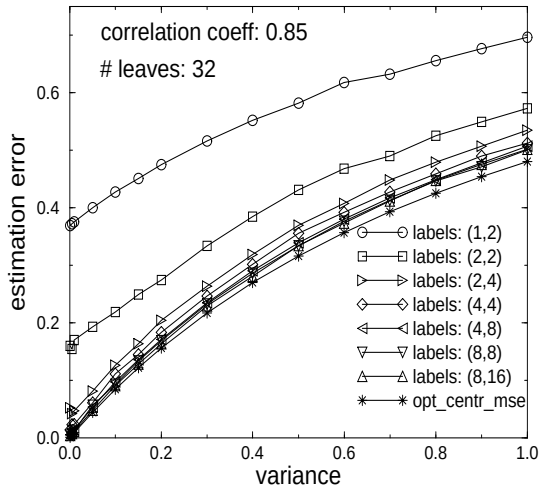


Fig. 2. Behavior of the system for several number of labels for quantizers (Q_1, Q_2) ($\rho = 0.85$, breakpoint initialization)

image and speech processing, and wireless communication. Yaacov Yesha was a program vice chair for the Seventh International Conference on Parallel and Distributed Computing Systems, 1994. He was a member of the program committees of the First International Conference on Computer Communications and Networks, 1992, and the 6th ISCA International Conference on Parallel and Distributed Computing Systems, 1993, and a session chair in the Advances in Digital Libraries forum, 1994. He is an IEEE member.