

Optimization of Decentralized Quantizers in Rate Constrained Data Fusion Systems

Maurizio Magarini and Arnaldo Spalvieri

Dipartimento di Elettronica e Informazione, Politecnico di Milano,

Piazza Leonardo da Vinci 32, 20133 Milano Italy.

Phone: +39-02-2399-3595 Fax: +39-02-2399-3413 E-mail: spalvier@elet.polimi.it.

ABSTRACT

A common model for distributed detection systems is that of several separated sensors each of which measures some observable, quantizes it, and communicates to a fusion center the quantized observation. The fusion center collects the quantized observations and takes the decision. This paper deals with the design of the quantizers when the channels between the sensors and the fusion center are subject to capacity constraints. The system of interest allows soft nonbreakpoint quantizers and nonindependent observations. We investigate two optimization techniques. The first technique finds a local minimum of the average misclassification risk by alternate optimization. The computational cost of the technique is exponential in the sum of the rates of the quantizers. To overcome this difficulty, a second technique based on a neural approach is presented, where a local minimum of the average misclassification risk is found by a stochastic approximation method.

1 INTRODUCTION

Distributed detection systems have received large attention in the past two decades, as documented in the special issue of the *Proceedings of the IEEE* [1]. A common model for these systems is that of several separated sensors each of which measures some observable, quantizes it, and communicates to a fusion center the quantized observation. The fusion center collects the quantized observations and takes the decision. Since the rate of transmission between the sensors and the fusion center is a cost, fine quantization of data may be not allowed. A crucial problem is therefore the design of coarse quantizers that satisfy a rate constraint and that introduce low degradation in the detection capability of the system. Tsitsiklis and Athans have shown in [2] that, when conditional independence of the observations given the hypothesis cannot be assumed, the design problem is N-P complete. Hence one is lead to renounce to global optimality and to study suboptimal strategies. Joint design of soft (multi-bit) quantizers has been studied by Longo *et al* in [3], where an alternate optimization technique is proposed. Specifically, the approach in [3] is to maximize the Bhattacharyya distance between the multivariate conditional probabilities of quantized data given the hypotheses. The potential weakness of this approach is that the Bhattacharyya distance is not the natural measure of performance of detection systems. Alternate optimization of the average misclassification risk was studied in [4]. Unfortunately, the computational complexity of the alternate optimization is exponential in the sum of the rates of the quantizers. To overcome this difficulty, here we propose a neural approach, where the boundary of

each quantization region is optimized by a stochastic approximation method. The results show that the performance of the neural method is slightly worse than the performance of the alternate optimization, and that the performance loss depends on the shape of the conditional probability density of the two hypotheses.

2 SYSTEM MODEL AND PROBLEM STATEMENT

For the sake of simplicity consider two scalar observations and binary detection. Extensions are straightforward. Let x_1, x_2 denote the observations, and assume that they are drawn from the continuous spaces $\mathcal{X}_1, \mathcal{X}_2$. In the classical formulation of the detection problem, a hidden discrete random variable (the *class*, or the *hypothesis*) is drawn together with the observation vector according to some known joint probability distribution. We call such a discrete random variable $c \in \mathcal{C} = \{c_1, c_2\}$. The goal of the detection system is to guess the hidden class given the observation vector. The decentralized detection system is modeled as a decision rule made by two scalar quantizers and a fusion center. Each scalar quantizer is allowed here to be a nonbreakpoint one. Quantizer $Q_n(x_n)$, $n = 1, 2$, is modeled as a mapping from \mathcal{X}_n to \mathcal{I}_n , where $\mathcal{I}_n = \{0, 1, \dots, I_n - 1\}$. Of course, the rate R_n of the n th quantizer is $R_n = \log_2 I_n$. Inversion of $Q_n(x)$ is hereafter intended as $Q_n^{-1}(i) = \{x_n \in \mathcal{X}_n : Q_n(x_n) = i\}$. The decision function performed by the fusion center, denoted $\mathcal{F}(i_1, i_2)$, is a mapping from $\mathcal{I}_1 \times \mathcal{I}_2$ to \mathcal{C} . The decision rule of the decentralized detection system, denoted $\mathcal{F}(x_1, x_2)$, is a mapping from $\mathcal{X}_1 \times \mathcal{X}_2$ to \mathcal{C} . As in [3], we assume that the processing to be performed at the fusion center is unlimited in complexity. In practice, this means that the fusion center is a look-up table with $2^{R_1+R_2}$ entries. A pictorial example of the decision rule for a specific two-dimensional decentralized detection system is illustrated in Fig. 1. The Bayesian risk (or cost) in deciding in favor of class $\hat{c} \in \mathcal{C}$ when x_1, x_2 is observed is

$$R(\hat{c}|x_1, x_2) = \sum_{i=1}^2 b(c_i \mapsto \hat{c})P(c_i|x_1, x_2), \quad (1)$$

where $b(c_i \mapsto \hat{c}) \geq 0$ is the risk of deciding in favor of class \hat{c} when $c = c_i$, and the familiar notation is adopted for the conditional probability of the class given the observation. Let $\mathcal{F}(x_1, x_2)$ be a decision rule. The classification performance of \mathcal{F} is measured by the expectation of the local risk (1) over the observation space:

$$R(\mathcal{F}) = \int_{\mathcal{X}_1} \int_{\mathcal{X}_2} R(\mathcal{F}(x_1, x_2)|x_1, x_2)p(x_1, x_2)dx_1dx_2.$$

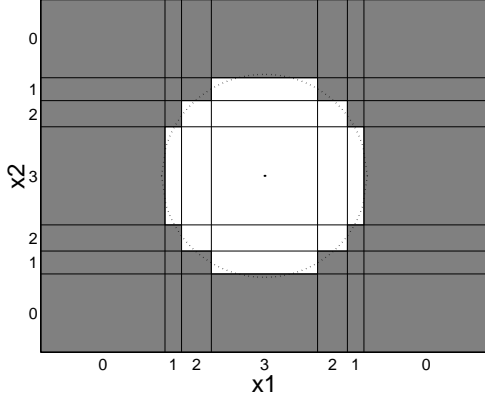


Figure 1: Example 2. Two-bit quantization ($\{0, 1, 2, 3\}$), $SNR = 10dB$, $r = 0$. Decision function (white=noise; gray=signal+noise). Dotted line: Bayes border.

where $p(x_1, x_2)$ is the (bivariate) probability density function of the observation. If we set $b(c_i \mapsto c_j) = 1$ for $i \neq j$ and $b(c_i \mapsto c_j) = 0$ for $i = j$, then the average misclassification risk (or, in short, risk) is equal to the average error probability. Therefore, from the Bayesian approach one comes to the following statement of the problem:

Fix $P(c_i)b(c_i \mapsto c_j) \forall i, j$, and find $\mathcal{F}(i_1, i_2)$, $Q_1(x_1)$, and $Q_2(x_2)$ that minimize $R(\mathcal{F}(Q_1(x_1), Q_2(x_2)))$.

The optimal decision rule, that will serve in the rest of the paper as a lower bound for the risk, is the Bayes test:

$$\mathcal{F}_B(x_1, x_2) = \arg \min_{\hat{c} \in \mathcal{C}} R(\hat{c} | x_1, x_2).$$

3 THE ALTERNATE OPTIMIZATION

The optimization algorithm is based on alternate optimization of the three functions (two quantizers and the decision rule). The optimality condition for Q_1 given Q_2 and \mathcal{F} is obtained from (1). Specifically, for all the points in \mathcal{X}_1 one finds the best index by taking the expectation of the local risk over \mathcal{X}_2 :

$$Q_1^{opt}(x_1) = \arg \min_{i_1 \in \mathcal{I}_1} \sum_{i_2=0}^{I_2-1} \int_{Q_2^{-1}(i_2)} R(\mathcal{F}(i_1, Q_2(x_2)) | x_1, x_2) \cdot p(x_2) dx_2, \forall x_1 \in \mathcal{X}_1. \quad (2)$$

A similar condition holds for Q_2 given Q_1 and \mathcal{F} .

The optimality condition for $\mathcal{F}(i_1, i_2)$ given Q_1 and Q_2 is obtained by taking the expectation of the local risk over the region indexed i_1, i_2 :

$$\mathcal{F}^{opt}(i_1, i_2) = \arg \min_{c \in \mathcal{C}} \int_{Q_1^{-1}(i_1)} \int_{Q_2^{-1}(i_2)} R(c | x_1, x_2) \cdot p(x_1, x_2) dx_1 dx_2, \forall i_1, i_2 \in \mathcal{I}_1 \times \mathcal{I}_2. \quad (3)$$

In the alternate optimization, we look for a new \mathcal{F} only when both quantizers are optimal for the old \mathcal{F} . Hence, after a proper initialization, the alternate optimization proceeds as follows:

1) determine Q_1 by (2),

2) determine Q_2 by (2), if the new Q_2 is different from the old Q_2 then go to 1,

3) determine \mathcal{F} by (3), if the new \mathcal{F} is different from the old \mathcal{F} then go to 1, else STOP.

Since each of the optimization steps does not increase the average risk, the procedure will terminate in a local optimum.

It should be noted that the practice of the procedure is limited by the complexity of calculation of the optimality conditions and by the size of the look-up table: both are proportional to $2^{R_1+R_2}$. Therefore the method can be actually applied only to coarse quantization of low-dimensional spaces. In the computer implementation, as well as in many actual detection systems, where continuous signals are digitally processed after analog-to-digital conversion, the continuous spaces \mathcal{X}_1 and \mathcal{X}_2 are discretized. In our experiments \mathcal{X}_1 and \mathcal{X}_2 are discretized in 256 slices of equal size. We have verified that such a slicing has negligible impact on system performance. Dealing with discrete signals, the integrals appearing in (2), (3) are replaced by sums in an obvious way.

4 THE STOCHASTIC APPROXIMATION

Let $\mathcal{P}_n = \{p_n(1), p_n(2), \dots, p_n(M_n)\}$, $M_n \geq I_n - 1$, $|p_n(i)| < \infty$, $p_n(i) > p_n(i-1)$, $n = 1, 2$, denote a set of points in \mathcal{X}_n . The set \mathcal{P}_n defines a partition of \mathcal{X}_n in $M_n + 1$ disjoint regions, which form the set $\mathcal{V}_n = \{v_n(0), v_n(1), \dots, v_n(M_n)\}$. Hence the set \mathcal{P}_n defines a mapping from points of \mathcal{X}_n to elements of \mathcal{V}_n . A successive mapping from \mathcal{V}_n to \mathcal{I}_n defines the possibly nonbreakpoint quantizer $Q_n(x_n)$. The set \mathcal{P}_n is the object optimized by stochastic approximation, while the mapping \mathcal{G}_n from \mathcal{V}_n to \mathcal{I}_n and the decision rule $\mathcal{F}(i_1, i_2)$ are kept fixed after a proper initialization. The stochastic approximation is based on a long (ideally infinite) random sequence $\{(\mathbf{x}^{(1)}, c^{(1)}), (\mathbf{x}^{(2)}, c^{(2)}), \dots\}$, $\mathbf{x} = \mathbf{x}_1, \mathbf{x}_2$, generated according to $P(c | x_1, x_2)p(x_1, x_2)$. Each sample of the sequence induces a stochastic estimate of the gradient of the risk with respect to the set of points, and the set of points is optimized according to the stochastic gradient algorithm. Since the conventional sample estimate of the risk is nondifferentiable, the classical Robbins-Monro method cannot be adopted. This problem can be circumvented by a technique similar to that proposed in [5], leading to the iterative algorithm

$$\begin{aligned} & \text{if } -d^{(k)} < x_1^{(k)} - p_1^{(k)}(i) \leq d^{(k)} \text{ then} \\ & p_1^{(k)}(i) = p_1^{(k-1)}(i) - \frac{h^{(k)}}{2d^{(k)}} (b(c^{(k)} \mapsto \mathcal{F}(\mathcal{G}_1(v_1^{(k)}(i)), Q_2^{(k)}(x_2^{(k)}))) \\ & \quad - b(c^{(k)} \mapsto \mathcal{F}(\mathcal{G}_1(v_1^{(k)}(i-1)), Q_2^{(k)}(x_2^{(k)})))) \\ & \text{else } p_1^{(k)}(i) = p_1^{(k-1)}(i), p_1(i) \in \mathcal{P}_1, k = 1, 2, \dots \end{aligned}$$

A similar algorithm holds for the points belonging to \mathcal{P}_2 . It is guaranteed that the algorithm terminates in a local minimum of the risk if the sequences of positive scalars $h^{(k)}$ and $d^{(k)}$ decay to zero according to the Kiefer-Wolfowitz conditions [6, §9]:

$$\sum_{k=1}^{\infty} h^{(k)} = \infty, \quad \sum_{k=1}^{\infty} h^{(k)} d^{(k)} \leq \infty, \quad \sum_{k=1}^{\infty} (h^{(k)} / d^{(k)})^2 \leq \infty.$$

5 SIMULATION RESULTS

The major trouble that we encountered in the experimental part of this work is that of local minima. When a cost function with multiple minima is considered, the results of both the alternate optimization and the stochastic approximation are strongly influenced by the initial guess. Due to space limitation, we cannot describe in detail the method used for determining the initial guess. The interested reader finds a detailed description of the method used for the alternate optimization in [4], and a similar approach has also been adopted for the stochastic approximation. The examples of interest are the same considered in [3]. We assume $P(c_1) = P(c_2) = 0.5$, $b(c_1 \mapsto c_1) = b(c_2 \mapsto c_2) = 0$, $b(c_2 \mapsto c_1) = \min\{1/t, 1\}$, $b(c_1 \mapsto c_2) = \min\{t, 1\}$. The results are presented as average risk against t .

Example 1: Known Signal in Spatially Correlated Noise
The observation model is:

$$\begin{aligned} c_1 &: \mathbf{x} = \mathbf{n} \\ c_2 &: \mathbf{x} = \mathbf{a} + \mathbf{n} \end{aligned}$$

where \mathbf{a} is the known signal vector and \mathbf{n} is a zero mean Gaussian noise vector with covariance matrix (let r be the spatial correlation coefficient):

$$\mathbf{S} = \begin{pmatrix} s^2 & rs^2 \\ rs^2 & s^2 \end{pmatrix}, \quad |r| < 1. \quad (4)$$

As in [3], the experiment concerns signals having equal energy, say $|a_1|^2 = |a_2|^2 = \mathcal{E}$, and per-channel signal to noise ratio $SNR = \mathcal{E}/s^2$. Fig. 2, reports the performance of the alternate optimization (AO), of the stochastic approximation (SA), and of the Bayes test (BT). It should be noted that, among the results that we worked out for this example, we have chosen to present the one where the difference between the two methods is higher.

Example 2: Spatially Correlated Unknown Signal in Uncorrelated Noise
The observation model is:

$$\begin{aligned} c_1 &: \mathbf{x} = \mathbf{w} \\ c_2 &: \mathbf{x} = \mathbf{s} + \mathbf{w}, \end{aligned}$$

where \mathbf{w} is a zero mean Gaussian noise vector with independent, unitary variance components and \mathbf{s} is a Gaussian signal vector with covariance matrix (4) where, now, r represents the spatial correlation coefficient of the signal and s^2 is the per-channel SNR . Fig. 3 reports the performance of the alternate optimization (AO), of the stochastic approximation (SA), and of the Bayes test (BT). In this case, as well as in many other instances of this example, the performance of the two methods is virtually the same.

REFERENCES

- [1] "Special issue on data fusion," IEEE Proceedings, Vol. 85, January 1997.
- [2] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized detection making and detection problems," IEEE Trans. on Automatic Control, Vol. AC-30, pp. 440-446, May 1985.

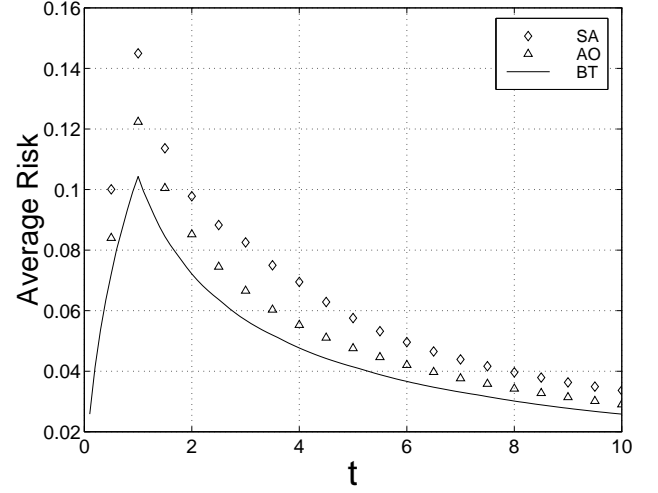


Figure 2: Example 1. $SNR = -5$ dB, correlation coefficient $r = 0.9$; $R = 3$ bits/sample for each quantizer.

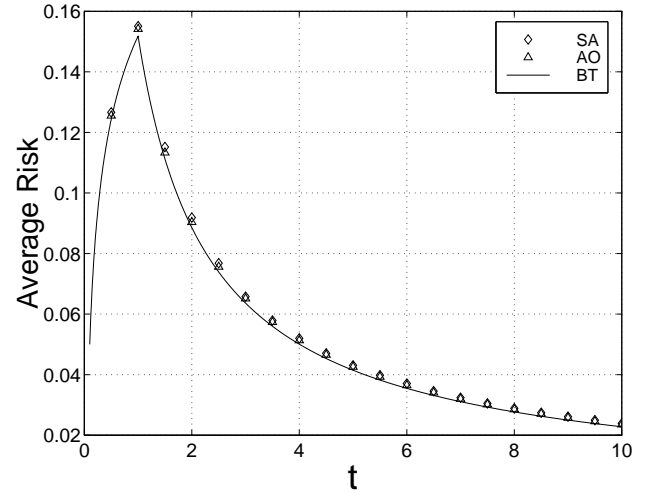


Figure 3: Example 2. $SNR = 10$ dB, correlation coefficient $r = 0.5$, $R = 2$ bits/sample for each quantizer.

- [3] M. Longo, T.D. Lookbaugh, and R. M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," IEEE Trans. on Information Theory, Vol. IT-36, pp. 241-255, March 1990.
- [4] M. Magarini and A. Spalvieri "Optimization of distributed detection systems under the minimum average misclassification risk criterion," IEEE Trans. on Information Theory, in press.
- [5] C. Diamantini and A. Spalvieri "Quantizing for minimum average misclassification risk," IEEE Trans. on Neural Networks, Vol. 9, pp.174-182, January 1998.
- [6] J. M. Mendel and K. S. Fu, "Adaptive, learning and pattern recognition systems," Academic Press, New York, 1970.