Fast Multilevel Quantization for Distributed Detection Based on Gaussian Approximation

Gökhan Gül, Michael Baßler
Fraunhofer Institute for Microengineering and Microsystems (IMM)
Carl-Zeiss-Str. 18-20, 55129 Mainz, Germany

Abstract—An iterative algorithm is derived for multilevel quantization of sensor observations in distributed sensor networks, where each sensor transmits a summary of its observation to the fusion center and the fusion center makes the final decision. The proposed scheme is composed of a person-by-person optimum quantization at each sensor and a Gaussian approximation to the distribution of the test statistic at the fusion center. The complexity of the algorithm is linear both for identically and non-identically distributed independent sensors. Experimental results indicate that the proposed scheme is promising in comparison to the current state-of-the-art.

Index Terms—Distributed detection, quantization, multisensor systems, cognitive radio, wireless sensor networks, signal detection

I. INTRODUCTION

Distributed detection has been the topic of various engineering applications such as fault detection for machine condition monitoring [1] or spectrum sensing for cognitive radios [2]. In parallel sensor networks each sensor makes an observation about a certain phenomenon, e.g. existence or absence of a signal, which is then transmitted to the fusion center for a final decision. Such a transmission over wireless channels—called centralized detection—is, however, impractical due to bandwidth and energy limitations [3]. Therefore, quantized versions of observations are transmitted to the fusion center—called decentralized detection—subject to the constraint that the loss of detection performance due to quantization is insignificant [4].

Considering the Bayesian cost function, decentralized detection was first studied by Tenney and Sandell for binary (two-level) sensor quantization without considering the design of data fusion algorithms [5]. This work was later extended by considering both the decision and the fusion rules employing numerical algorithms that find person-by-person optimal solutions [6]. Two level quantization studied by [5], [6], were also generalized to multilevel quantization, for instance by considering fuzzy membership functions [7].

In order to be able to provide more elegant solutions, which are not limited to only a few sensors [5] or only binary decisions [6], pseudo objective functions based on distances between probability distributions are adopted. In [8] the J-divergence was shown to improve the performance of binary quantization by including additional quantization levels. Intuitively, it may be thought that the means considered in the J-divergence do not carry sufficient information and the objective function should also include the variances of the distributions. In [9], this idea was applied by considering the deflection distance. In a more recent work [10], the J-divergence and the Bhattacharyya distance are considered separately for the optimization of the parallel sensor network, improving the Lee and Chao’s J-divergence based method [8].

Chernoff information (CI) is known to have nice asymptotical properties [11] and maximizing the CI amounts to minimizing the upper bound on the true objective function [12]. Based on this idea, CI was used for multilevel quantization in [11]. In a further study scalable solutions were obtained by locally maximizing the Chernoff information at each sensor, but no algorithmic solution was proposed [13]. Recently, Chernoff information and deflection distance were used for the locally optimum quantization of sensor observations, presenting explicit algorithmic solutions [14], [15].

Although considering variances along with means is a valuable idea, its implementation via deflection distance has not been successful in comparison to the Chernoff distance [15]. The reason may lie in the fact that the deflection distance is only a very special case of the optimization process with the Chernoff distance and it is ill-conditioned if the variance is very small. In this paper, both the means and the variances are considered employing a Gaussian approximation to the overall distributions of the test statistic under each hypothesis. This approximation is motivated by the central limit theorem, which implies that the approximation is accurate if the total number of sensors is large enough. Together with a person-by-person optimization process for the sensor decisions, exponential computational complexity of the original optimization problem is reduced to a linear time complexity. The algorithm is iterative in nature and achieves near optimal solutions. Numerical results illustrate the feasibility of the proposed algorithm in comparison to the recent literature [14], [15].

The rest of this paper is organized as follows. In Section II, the decentralized detection problem is introduced. In Section III, the proposed algorithm is derived and its computational complexity is analyzed. In Section IV, the performance of the proposed algorithm is evaluated and compared to [14], [15]. Finally in Section V, the paper is concluded.

II. DECENTRALIZED DETECTION

Consider a distributed detection network with \( K \) decision makers \( \phi_1, \ldots, \phi_K \) and a fusion center \( \gamma \) as illustrated by...
Figure 1. Each sensor \( \phi_k \) makes an observation \( y_k \in \Omega_k \) from a certain phenomenon, where \( \Omega_k \) is an interval, and gives a multilevel decision \( u_k \in \{0, \ldots, N_k - 1\} \). The phenomenon is modeled by a binary hypothesis testing problem

\[
\begin{align*}
\mathcal{H}_0 : Y_k &\sim f^k_0 \\
\mathcal{H}_1 : Y_k &\sim f^k_1
\end{align*}
\]

(1)

where the random variables (r.v.s) \( Y_k \) corresponding to the observations \( y_k \) are mutually independent and follow the probability density functions \( f^k_0 \) or \( f^k_1 \), conditioned on the hypothesis \( \mathcal{H}_0 \) or \( \mathcal{H}_1 \). The fusion center receives multilevel decisions from all sensors and gives a binary decision \( u_0 \).

Optimum quantization, which minimizes the error probability of the fusion center is known to be the monotone likelihood ratio test [4]. Assuming that the likelihood ratio function \( l_k = f^k_1/f^k_0 \) is strictly monotone, thresholding can be done directly over the observations. Thus, the decisions can be obtained by

\[
\phi_k(y_k) = u^i_k \quad \text{if} \quad \lambda^{i_k}_{k-1} \leq y_k < \lambda^{i_k}_k
\]

(2)

where \( \lambda^i_k \) denotes the thresholds, \( k \in \{1, \ldots, K\} \) denotes the indices of sensors and \( i_k \in \{1, \ldots, N_k\} \) denotes the indices of the multilevel decision \( u_k \) for the \( k \)th sensor. The upper and lower thresholds are given by \( \lambda^0_k := \inf \Omega_k \) and \( \lambda^{N_k}_k := \sup \Omega_k \), leaving \( N_k - 1 \) unknown thresholds to be determined per sensor. From (1) and (2) the probability mass functions of the decisions conditioned on the hypothesis \( \mathcal{H}_m \), \( m \in \{0, 1\} \), can be found by

\[
p^k_m(u^k_k) = \int_{\lambda^{i_k}_{k-1}}^{\lambda^{i_k}_k} f^k_m(y)dy.
\]

(3)

Let \( p_0 \) and \( p_1 \) denote the joint probability mass functions of the random variables \( U_k \), corresponding to the multilevel decisions \( u_k \), conditioned on the hypotheses \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \), respectively.

Furthermore, let the transmitted decisions \( u_k \) be reformed by the fusion center optimally as

\[
u_k := \log \frac{p^k_1(u_k)}{p^k_0(u_k)}.
\]

Then, the optimum test at the fusion center can be obtained by

\[
\log \frac{p_1(u_1, \ldots, u_K)}{p_0(u_1, \ldots, u_K)} = \sum_{k=1}^K \log \frac{p^k_1(u_k)}{p^k_0(u_k)} = \sum_{k=1}^K \frac{\mathcal{H}_1}{\mathcal{H}_0} u_k \geq \lambda_0, \quad (4)
\]

where \( \lambda_0 \in \mathbb{R} \) is a suitable threshold. Since test statistic in (4) corresponds to the summation of \( K \) random variables \( U_k \), the probability mass function of the sum can be obtained by \( K \)-fold convolution of the marginal mass functions as

\[
g_m(z) = \sum_{i_1=1}^{N_1} \cdots \sum_{i_K=1}^{N_K} p^1_{m}(u^{i_1}_1) \cdots p^K_{m}(u^{i_K}_K) \delta \left( \sum_{k=1}^K u^{i_k}_k - z \right),
\]

where \( \delta \) is the dirac delta function. Then, for \( N = \dim(g_m) \) the minimum error probability can be found by

\[
P_E = \sum_{n=1}^N \min \left( g_0(n), g_1(n) \right),
\]

using an order preserving bijective mapping \( z \mapsto n \). Let all non-trivial thresholds of the sensors be represented by the set of parameters

\[
\lambda = \{ \lambda^{i_k}_k : i_k \in \{1, \ldots, N_k - 1\}, k \in \{1, \ldots, K\} \}.
\]

Furthermore, let \( p^\lambda_m \) denote the joint mass defined explicitly by its dependence on \( \lambda \). Then, the overall optimization problem can be formulated as

\[
P_E = \min_{\lambda} \sum_{n=1}^N \min \left( g_0(n; p^\lambda_m), g_1(n; p^\lambda_m) \right).
\]

(5)

The problem in (5) is of exponential complexity even for identically distributed sensor observations, since there are counterexamples showing that identical sensor decisions are not always optimum for identically distributed sensors. [16]. However, according to central limit theorem the probability mass function \( g_m \) can be well approximated by a Gaussian probability density function \( \tilde{g}_m \) with mean \( \mu_m \) and the variance \( \sigma^2_m \) for large enough \( K \), if \( U_k \) are identically distributed. Hence, the optimization problem in (5) can also be well approximated as

\[
P_E = \min_{\lambda} \int_{\mathbb{R}} \min \left( \tilde{g}_0(y; \mu_0(\lambda), \sigma^2_0(\lambda)), \tilde{g}_1(y; \mu_1(\lambda), \sigma^2_1(\lambda)) \right) dy.
\]

(6)

III. OPTIMIZATION OF THE SENSOR NETWORK

In this section an iterative algorithm will be derived for the solution of (6). The algorithm is capable of quantizing both identically as well as non-identically distributed sensor observations.
A. Derivation of the Algorithm

First, it can been seen that using the identity
\[ \min(a, b) = \frac{a + b}{2} - \frac{|a - b|}{2} \]
the optimization problem in (6) can be written as
\[
P_E = \min_{\lambda} \left( \frac{1}{2} \int_{\mathbb{R}} \left[ \tilde{g}_0(y; \mu_0(\lambda), \sigma_0^2(\lambda)) + \tilde{g}_1(y; \mu_1(\lambda), \sigma_1^2(\lambda)) \right] \right.
\]
\[-\left| \tilde{g}_0(y; \mu_0(\lambda), \sigma_0^2(\lambda)) - \tilde{g}_1(y; \mu_1(\lambda), \sigma_1^2(\lambda)) \right| dy, \]

Since the integral of the two integrands results in 2, we have
\[
P_E = \min_{\lambda} \left( \frac{1}{2} \int_{\mathbb{R}} \left[ \tilde{g}_0(y; \mu_0(\lambda), \sigma_0^2(\lambda)) - \tilde{g}_1(y; \mu_1(\lambda), \sigma_1^2(\lambda)) \right] |dy, \]
which can be simplified to
\[
P_E = \frac{1}{2} \max_{\lambda} \int_{\mathbb{R}} \left[ \tilde{g}_0(y; \mu_0(\lambda), \sigma_0^2(\lambda)) - \tilde{g}_1(y; \mu_1(\lambda), \sigma_1^2(\lambda)) \right] |dy.
\]

Next, by separating the integration domain, we obtain
\[
P_E = \frac{1}{2} \max_{\lambda} \int_{\tilde{g}_0(\lambda) > \tilde{g}_1(\lambda)} \left[ \tilde{g}_0(y; \mu_0(\lambda), \sigma_0^2(\lambda)) - \tilde{g}_1(y; \mu_1(\lambda), \sigma_1^2(\lambda)) \right] dy
\]
\[+ \int_{\tilde{g}_0(\lambda) \geq \tilde{g}_1(\lambda)} \left[ \tilde{g}_1(y; \mu_1(\lambda), \sigma_1^2(\lambda)) - \tilde{g}_0(y; \mu_0(\lambda), \sigma_0^2(\lambda)) \right] dy.
\]

From \( \int_{\tilde{g}_0(\lambda) > \tilde{g}_1(\lambda)} \tilde{g}_m(y)dy = 1 - \int_{\tilde{g}_1(\lambda) \geq \tilde{g}_0(\lambda)} \tilde{g}_m(y)dy \) we further get
\[
P_E = \max_{\lambda} \int_{\tilde{g}_1(\lambda) \geq \tilde{g}_0(\lambda)} \left[ \tilde{g}_1(y; \mu_1(\lambda), \sigma_1^2(\lambda)) - \tilde{g}_0(y; \mu_0(\lambda), \sigma_0^2(\lambda)) \right] dy.
\]

Finally, by considering (4) we can write (7) as
\[
P_E = \max_{\lambda} \int_{\lambda_0(\lambda)}^{\infty} \left[ \tilde{g}_1(y; \mu_1(\lambda), \sigma_1^2(\lambda)) - \tilde{g}_0(y; \mu_0(\lambda), \sigma_0^2(\lambda)) \right] dy,
\]
where \( \mu_1 > \mu_0 \). Next, consider the following assumption.

**Assumption III.1.** All thresholds except for the threshold \( \lambda_{k}^{i_k} \) are known in solving the optimization problem in (8).

Using Assumption III.1 the multivariable optimization problem given by (8) can be reduced to
\[
P_E = \max_{\lambda_{k}^{i_k}} \int_{\lambda_0(\lambda_{k}^{i_k})}^{\infty} \left[ \tilde{g}_1(y; \mu_1(\lambda_{k}^{i_k}), \sigma_1^2(\lambda_{k}^{i_k})) - \tilde{g}_0(y; \mu_0(\lambda_{k}^{i_k}), \sigma_0^2(\lambda_{k}^{i_k})) \right] dy.
\]

Dropping the dependencies of \( \mu_m, \sigma_m \) and \( \lambda_0 \) to \( \lambda_{k}^{i_k} \), and taking the integral, the \( P_E \) minimizing threshold can eventually be obtained by
\[
\lambda_{k}^{i_k} = \arg \max_{\lambda_{k}^{i_k}} \text{erf} \left( \frac{\mu_1 - \lambda_0}{\sigma_1} \right) - \text{erf} \left( \frac{\mu_0 - \lambda_0}{\sigma_0} \right), \quad \text{where the fusion threshold}
\]
\[
\lambda_0 = \frac{\mu_1 \sigma_0^2 - \mu_0 \sigma_1^2}{\sigma_0^2 - \sigma_1^2}
\]
\[= \sigma_0 \sigma_1 \sqrt{(\mu_0 - \mu_1)^2 \log(\sigma_0/\sigma_1) + (\sigma_0^2 - \sigma_1^2)} \quad \text{(10)}
\]
is derived by solving
\[
\tilde{g}_0(y; \mu_0, \sigma_0^2) = \tilde{g}_1(y; \mu_1, \sigma_1^2)
\]
for \( y \). It can be seen that the optimization problem in (9) depends only on \( \mu_m \) and \( \sigma_m^2 \) by considering (10). Remembering that we also have \( u_{i_k}^k = \log p_k(u_{i_k}^k)/p_0(u_{i_k}^k) \), the overall means and variances of \( \tilde{g}_0 \) and \( \tilde{g}_1 \) can be found as
\[
\mu_m = \log \frac{p_k(u_{i_k}^k)}{p_0(u_{i_k}^k)} p_m(u_{i_k}^k) + \log \frac{p_k(u_{i_k}+1)}{p_0(u_{i_k}+1)} p_m(u_{i_k}+1) + c_{\mu_m}
\]
and
\[
\sigma_m^2 = \left( \log \frac{p_k(u_{i_k}^k)}{p_0(u_{i_k}^k)} - \mu_m \right)^2 p_m(u_{i_k}^k)
\]
\[+ \left( \log \frac{p_k(u_{i_k}+1)}{p_0(u_{i_k}+1)} - \mu_m \right)^2 p_m(u_{i_k}+1)
\]
\[+ c_0 \mu_m^2 + c_1 \mu_m + c_2 + c_{\sigma_m^2} \quad \text{(12)}
\]
By Assumption III.1, the first two summands of (11) and (12) are all dependent on the parameter \( \lambda_{k}^{i_k} \) through (3), including the (two related terms of the) mean of the \( k \)th sensor
\[
\mu_k = \sum_{i_k=1}^{N_k} \log \frac{p_k(u_{i_k}^k)}{p_0(u_{i_k}^k)} p_m(u_{i_k}^k),
\]
whereas all other terms including the other \( \mu_{m, s} \) are some constants, which can be calculated as
\[
c_{\mu_m} = c_3 - \sum_{1 \leq i_k \leq N_k} \log \frac{p_k(u_{i_k}^k)}{p_0(u_{i_k}^k)} p_m(u_{i_k}^k),
\]
\[c_{\sigma_m^2} = c_4 - \sigma_m^2 \quad \text{(13)}
\]

with together with
\[
c_0 = \sum_{1 \leq i_k \leq N_k} \log \frac{p_k(u_{i_k}^k)}{p_0(u_{i_k}^k)} p_m(u_{i_k}^k),
\]
\[c_2 = \sum_{1 \leq i_k \leq N_k} \log \frac{p_k(u_{i_k}^k)}{p_0(u_{i_k}^k)} p_m(u_{i_k}^k),
\]
\[c_3 = \sum_{k=1}^{K} \mu_m, \quad c_4 = \sum_{k=1}^{K} \sigma_m^2 \quad \text{(14)}
\]
where
\[
\sigma_m^2 = \sum_{i_k=1}^{N_k} \left( \log \frac{p_k(u_{i_k}^k)}{p_0(u_{i_k}^k)} - \mu_m \right)^2 p_m(u_{i_k}^k). \quad \text{(15)}
\]
Algorithm 1 Gaussian based multilevel observation quantizer

**Input:** $f_0, f_1,$ and initialized $\lambda_k^0, p_0^k, p_1^k$

**Output:** Updated thresholds $\lambda_k^h$ and probabilities $p_0^h, p_1^h$

1: do
2: Find $c_3$ and $c_4$ by using (14)
3: for each $k \in \{1, \ldots, K\}$ do
4: Calculate $c_{2m}^2$ by using (13)
5: for each $i_k \in \{1, \ldots, N_k\}$ do
6: Calculate $c_{\mu m}, c_0, c_1, c_2$ using (13) and (14)
7: Update (11) and (12) by the coefficients found
8: Update the threshold $\lambda_k^h$ by solving (9)
9: Update the probabilities $p_0^h, p_1^h$ by using (3)
10: end for
11: end for
12: while $\forall \lambda_k^h$ converges

### B. Complexity Analysis

The proposed scheme which iteratively computes the optimum quantization thresholds is given by Algorithm 1. In every iteration computation of the sum of the means $c_3$ and variances $c_4$ requires $O(K)$ computations, where $O(\cdot)$ is the standard Landau notation. Given $c_4$, calculating $c_{2m}^2$ by (13) and (15) requires $O(N_k)$ computations for each $k$. The rest of the coefficients, i.e., $c_0, c_1, c_2$, and $c_{\mu m}$ given $c_3$ for each $k$ and $i_k$ requires no more than $O(N_k)$ computations per coefficient.

Solving an equation in one parameter can be performed using iterative methods such as the Nelder-Mead method [17]. Since the complexity of equation solving is independent of the total number of sensors, order of computations for Algorithm 1 is linear in $K$. In comparison to [15], the proposed scheme may be preferable since the computational complexity of equation solving in [15] is linearly increasing with $K$. This raises the overall complexity of CI based quantization to $O(K^2)$.

### IV. Numerical Results

In this section, performance and convergence properties of the proposed algorithm are evaluated over identically as well as non-identically distributed independent sensor observations. In all simulations firstly $\lambda_k^h$ are initialized uniformly on $[\lambda_k^0, \lambda_k^{N_k}]$. Then, Algorithm 1 is run until the sum of the absolute values of the differences of thresholds from one iteration to the next is less than a common threshold, e.g. 0.001.

#### A. Identically distributed observations

Consider the uniform vs. linear (UL) and Gaussian vs. Gaussian (GG) hypothesis testing problems,

$$H_m : Y_k \sim f_m(y_k) = \frac{1}{2} y^m \mathbf{1}_{\{0 \leq y \leq 2\}}(y), \quad m \in \{0, 1\},$$

$$H_m : Y_k \sim f_m(y_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y - m)^2}{2\sigma^2}}, \quad m \in \{0, 1\},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. First, let us consider the UL-problem and suppose that each sensor transmits 3-bits to the fusion center. Convergence of the optimum thresholds of the Gaussian method in comparison to Chernoff based quantizer is plotted over various number of sensors in Figure 2. Notice that the thresholds of Gaussian quantizer do not converge to those of Chernoff based quantizer asymptotically. Now let us consider the UL- and GG-problems with 1-, 2- and 3-bit quantizations for various total number of sensors. Figures 3 and 4 illustrate the results of these experiments, respectively, relative to the centralized detection. In Figure 3, the performances of the Gaussian method for the asymptotic cases are also given by dashed lines. In both figures, the $P_E$ values are determined by extensive Monte-Carlo simulations.

It can be seen that considering the proposed scheme is useful for 1-bit quantization, whereas the results are almost the same for all other cases.

#### B. Non-identically distributed observations

Consider the hypothesis testing problem with non-identically $\chi^2$ distributed random variables,

$$H_m : Y_k \sim f_m(y_k) = \frac{W}{2^{W/2}} (\psi_k + 1)^{\frac{m}{2}} \Gamma \left( \frac{W}{2} \right), \quad m \in \{0, 1\},$$

where $W$ is the number of samples collected by each sensor, $\psi_k$ is the signal-to-noise ratio (SNR) and $\Gamma$ is the gamma function. These distributions arise from a signal detection problem, where each detector is an energy detector over a static channel model facing a presumably different SNR. The details of this problem can be found in [15, p. 42]. We assume that each sensor collects $W = 10$ samples, as in [15], and the SNR range of $[-3, 2]$ dB is divided uniformly to the total number of sensors in the network as in [13]. The error probability $P_E$ is evaluated with extensive Monte-Carlo simulations as before. Figure 5 illustrates the outcome of this experiment. The proposed scheme is slightly outperformed by Chernoff based quantizer for 2- and 3-bit quantization and the results are almost identical for 1-bit quantization.

### V. Conclusion

An algorithm was derived for the optimization of sensor networks which consist of a finite number of sensors and a fusion center. The algorithm is capable of quantizing both identically as well as non-identically distributed independent...
sensor observations. The proposed scheme reduces the exponential complexity of fusion by making Gaussian approximation to the conditional distributions of the test statistic and the exponential complexity of quantization at sensor nodes by employing person-by-person optimization. This results in a computational complexity which is linear in total number of sensors. The motivation behind the proposed approximation is due to central limit theorem, which asserts that the accuracy of approximation improves as the number of sensors increases. Numerical results indicate feasibility of the proposed scheme in comparison to the state-of-the-art. Since both schemes result in different thresholds, even for the same detection accuracy, different false alarm probabilities may be preferable in different applications. Linear complexity of the algorithm makes it also promising in comparison to the state-of-the-art, which has a quadratic complexity. The proposed method can further be improved by generalizing the Gaussian approximation to a set of distributions parameterized by a certain parameter and performing the optimization over both the related parameter as well as the thresholds. Moreover, the method can easily be extended to density functions which have non-monotone likelihood ratios by using generalized inverse functions.

REFERENCES


